

XII. THE REVIEW PANEL'S ASSESSMENT

A. INTRODUCTION

The preceeding four chapters contain the views of economics scholars whose own research has been focused on the development of the CVM; their interests and expertise in (with) the method was reflected in our repeated references in Part I to their earlier works.

As stated in Chapter I, the breadth of our assessment of the CVM is greatly enhanced by looking also to outstanding scholars whose research interests are a step removed from CVM research for their assessments of the state of the arts of the method. Thus, our Review Panel, consisting of Professors Arrow, Kahneman, Rosen and Smith, offer the Comments given in Sections B - E in response to, first their pre-Conference reading of Tart I and secondly, the Conference presentations of Professors Randall, Bishop, Heberlein, Freeman and V. Kerry Smith.

B. COMMENTS BY PROFESSOR KENNETH ARROW

The fundamental question being raised by the CVM approach (but not confined to it) is the transferability of results from one realm of observation -- observation of human behavior -- into another realm.

For various welfare reasons, we agree that a certain kind of pseudo-pricing will be, if demonstrably accurate, a useful basis for deciding on certain public goods measures, environmental measures, or whatever. We have a set of observations that don't relate to that field. We want to use these other observations, in this case responses to verbal criteria -- in other words a different kind of behavior -- and transfer them. Now this occurs not only in the context of public goods. In fact, it occurs not only in the context of economics; psychologists are always making observations in the form of experiments as well as in the form of field observations in certain limited circumstances and extrapolating to make inferences to other circumstances. At least that is presumably the purpose of the inquiry. One is not seriously interested in the response of a few college students to waving little rewards in front of their faces. Presumably you are using questionnaires because you are learning something, let's say, about your subjects' resistance to new information; their ability to translate given conditions into certain actions, which is a little more fashionable today; or to learn about difficulties of communication, say restricted communication networks, and how they manifest themselves in certain behavior.

Unfortunately there does not seem to be any systematic methodology for transferring results of experimental, rather small scale, situations to other situations, more specifically to uncontrolled situations. wow probably this transition will never be done well anywhere. Since I misspent part of my life as a meteorologist, I am acquainted with the fact that knowing physics very well is only of mild usefulness for weather forecasting, and yet we know very well that the elementary principles which determine the weather are in fact governed by the laws of physics, and our knowledge there is far deeper than we have in psychology or economics. SO it is not surprising. that these transfers from one situation to another are difficult -- it is very typical.

One question is, does it mean anything at all? If you ask somebody a question you will get an answer. What this has to do with how much somebody really values something is conjectural. What kind of evidence do you bring to bear on this? One source of evidence is the consistency in the answers.

Actually, we generally do feel fairly safe for the most part (and psychologists certainly do, I think with some good reasons) in transferring the qualitative implications of their experiments. We learn that if people have taken a strong position it is not too easy to get them to change it, even in the presence of overwhelming information. This corresponds to the observations we make in real life, say, when we deal with our students. I don't know anybody who has made the attempt to say how much teaching will we need to overcome a given amount of a priori information.

One curious thing which was a subject of interest in psychology for a while and seems to have a lesson for us, is the work on scaling of subjective phenomena, particularly by S.S. Stevens and his students. He would, for example, play a couple of notes on the piano and then play a

third note and then ask "Is this closer to note A or note B?" The first time I ran across this on a doctoral examination I kept on asking the student what the question meant. From my ordinalist viewpoint, I couldn't imagine what it could possibly mean. Finally his professor said, "Anybody but an economist would understand that."

There was a reality there, To be sure, if you ask a question you get an answer. The reality was that you start again with two different reference notes. You get a scale which is a linear transformation of the original scale. That is a refutable hypothesis -- at least I was being assured of the fact that it was not refuted. It tested out very very well,

This meant that here was some reality. Unfortunately it isn't very clear sometimes whether that is the reality we are interested in for our purposes. I do notice that for whatever reasons that line of investigation seems not to have gone any further.

Now we do find a problem. Consider the structure of an ordinary demand curve. We have a lot of observations, let's say a cross-section comparison on prices and quantities, and we derive the demand curve. In Chapter VI of the Assessment Report, it is noted that, in deriving this demand curve, when you do something as simple as change your assumptions on the distribution of the residuals, you get wildly different elasticities. This points to the fact that, in assessing methods such as the CVM, the demand curve should not be considered as some kind of "reality" to which we should hope to aspire. As pointed out by the authors, demand curves themselves are problematic.

Consider a problem closer to the sort of things we are talking about (the CVM) -- a businessman who wants to produce a new product. He wants to know what he can sell it for. Of course there are questions of his costs, but that is in essence a private type of information that he or she can dispose of. What he or she has to look at is the worth. How much will the public pay for the product? Businessmen don't know, and more than fifty percent of all the new products put on the market fail. I don't mean fifty percent of ideas don't succeed, I mean fifty percent of the products which have already reached the point of market introduction are failures. So it is obvious that the estimation of the demand functions by businessmen is tinged with a large degree of error.

I'm trying to put some context on this question of what the CVM may really provide, how much one can expect from it. One more word on this subject -- I think this was brought up by one of the speakers -- about field experiments.

By considering contingent valuation as compared with other forms of indirect measurement, we have unduly limited the number of possible ways of getting information. There are others, and indeed field experiments -- though not quite parallel to these -- were, at least a few years ago, a major source of economic inquiry. The income-maintenance experiments, the health insurance experiment, the housing allowance experiments, were large-scale field experiments. These studies typically involved private goods, so the results we got from them had as much significance as one could possibly place on them, and should have been (at least in principle) a great deal more reliable than the observations made from uncontrolled observations -- the sort of thing you've been dealing with in CVM experiments. In fact, very interestingly, the results were not all that different from results obtained from earlier studies based on secondary data. And furthermore,

rather significant ranges of error were found in those field experiments concerning private goods; for example, a considerable range of error was found in the elasticity of the supply of secondary labor in the case of the income-maintenance experiment, depending on what was being controlled for, or what you were allowing to vary.

This suggests some basic research. Now that may be the last thing one wants to hear around here, given the emphasis earlier on the scarcity of research funds, but one possible line is to take a field where CVM is unnecessary. This is just the place to do the research. The reason is, of course, that this is the only way you will ever be able to calibrate your measures. Comparisons of the CVM with other methods with all the associated difficulties described in Chapter VI are extremely important. If you are finding, by two conceptually quite different methods, numbers that are the same or similar, then -- while you can't be quite sure that the reality that you are reaching is the reality that you want -- at least you are reassured that you are likely to be measuring something real. We were reassured in this way in Steven's work on scaling. He scaled by several different methods, some of which seemed totally improbable to an economist, and yet the results were consistent. So I think trying to reconstruct ordinary demand curves by survey methods as well as by field experiments seems the sort of thing that is needed to validate the CVM for that other rather large class of cases where CVM seems to be the only method that makes any sense, short of course, of sheer a priorism or guesswork.

There have been a lot of statements made on the matter of the "hypothetical elements," of the CVM and I would like to comment on several classes of what has been referred to as hypothetical bias. One problem is that the commodity in the CVM is hypothetical. Again, that is not as unique as it seems to be, because as indicated, every time there is a new product you have a hypothetical element in your story. There are questions. Whether the answers are guessed by the producer or by some kind of consumer inquiry is another matter. The fact is we are in a world in which there are new things, and this is not exceptional -- new products are constantly introduced in the market. In many industries, where we define the word "product" rather narrowly, fifty percent of the products sold at any time are less than three years old. In these industries there is always guessing about the receptivity of the market, and the guesswork is pretty clear from the fact that they fail every now and then. That we are dealing with hypothetical commodities is not so much a drawback as a fact.

I find the hypothetical bias concerning payment more serious than that about commodities. This is the concern of those who follow the economists' tradition which criticizes hypothetical questions. Verbal answers don't hurt the way cash payments do. Some evidence suggested that there was a real difference between cash payments and hypothetical payments. But on the whole the discrepancy was not as bad as one might fear.

Any time you have an irreversible element, especially one of some significance, you are changing the world, and the situation is hypothetical. It can never be put back. Now in the case of some environmental situations there is some chance for correction in the sense that there are similar situations in diverse geographies, so one can have a feedback process. If in retrospect it turns out you wished you hadn't made some change, you needn't make it elsewhere. This is the process which prevents blind investment from being totally disastrous -- that there are

enough similarities to be able to make an inference from one case to another. This reminds me that I haven't seen any discussions of cross-situation comparability, which is a way to get demand curves by essentially comparing situations at different times, and/or different places. It is not clear to me whether there has been enough attention paid to this. There's too much geographic specificity in the studies reported here.

Let me continue by discussing briefly some of the other biases addressed here. Neither the empirical evidence nor the theoretical arguments convinced me that strategic bias is liable to be significant. Sherwin Rosen does raise a point: Supposing I am asked, "From now on will you use the survey data?" That is, will survey data form the basis of our judgements? Then, indeed, I suppose one might have some problems. But let's not think that far ahead. This means the whole discussion about Vickery auctions and the like, which are basically incentive-compatibility methods, are really beside the point. I don't think this has much to do with the basic issue.

Several other biases were mentioned, and I will go over them very briefly. One was the vehicle bias. I must say, I didn't have a conviction from my reading that the vehicle bias does indeed matter. There is nothing irrational about a difference in responses in this case. If I'm going to finance a change by use permits, it is significantly different from the case where I finance it by general taxation. Let me put it differently -- it would be irrational if you did not get a difference in the responses in these two situations. It is a fact that WTP depends on who gets the "P," and on what that means. This is very reasonable in some circumstances. Now for others, it may not be. You can get the framing problem. Say you get two methods of payment where every individual in fact is paying the same amount, or at least his or her random expected payment is about the same. Then if the responses differ, you may have a real vehicle bias. But if it is merely that taxing according to one principle, like use permits, gives a different result entirely than putting a general price, for example a bonus tax, on the public at large, then I find nothing remarkable. I do not have the conviction that these two different sources have been well expressed.

We need to see more data than is usually supplied, because these distributions of willingness-to-pay were very skewed. The mean was always much higher than the median. If you have a highly skewed demand, so that few people have a high value for it, there are certainly implications for methods of financing. It certainly suggests that a method which captures the surplus by individuals, even though it may be inefficient in some technical sense, may be superior to an alternative which tries to distribute the cost, say, in some very broad way. It seems to me that the implication of this distribution is not that there is an error of measurement. Now, it may be, but I am assuming that it is not. It is a perfectly real possibility that some people value these things much more highly than others would -- visibility or the right to hunt or whatever. It does suggest that some method of benefit taxation is appropriate. There are such striking differences that averaging them out may be unfair and may have legitimate political repercussions.

Again, on the information bias topic, I found that several different strands seem to have been drawn together, some of which are not biases at all. There was a lot of reference to information about other people's

preferences. NOW, in some sense this is the last thing you want. If you are worried at all about strategic bias, then you do not want information about other people's preferences, because you make strategic bias easy to achieve and you may induce it by your method of response. There is another reason that you might be concerned with other people's preferences and that is second-hand information: "Now, if everybody else thinks it's a good idea, it probably is a good idea, and I know I am uninformed and other people know a good deal more about it." But that requires deliberate modelling to take that into account. It can't be done by simply adding up WTP's.

Other kinds of information seem to be proposed which are simply explaining the matter in greater detail, greater specificity. These are already connected with hypothetical bias with regard to commodities. They are simply trying to explain the commodity in greater detail. Someone who knows more about surveys than I do would be better able to evaluate just how much you can present, for example, before the difficulties in processing the information presented begin to outweigh the benefits from having more information. This is something that I assume something is *known about, with the many years of survey research in this country.

How you make a survey situation realistic is something I don't know. My impression is that the evidence indicates that the more you structure a situation to be a pseudo reality the more real-like are the results YOU elicit. But of course that usually has some price.

Finally, addressing the question of accuracy, there is an interesting question: What, even ideally, do we mean by accuracy? What is the reference? What is the reality to which we refer? We want to compare the outcome to some truth. Well, suppose we had infinite research resources, what would we mean? I suppose we want some kind of ex-post valuation -- even that, of course, is hypothetical. One trouble is that in economics, as well as in other social sciences, almost all economic reality has to do with counterfactuals. What do we mean by saying that you quote a price? Is this prize the cost of time commodity, or what you would give up to buy it? This is full of the subjunctive mood. This is not confined to economics, but economics has developed this logic. Almost everything, all the concepts of marginalism, are counterfactual statements. They are statements comparing something to what would be true if it were not so. "If you produce one unit less," or statements of that kind -- "if your income was one unit higher." There is a certain impalpable air of alternatives that are not being realized in some sense. sometimes, very occasionally, nature will supply you with that experiment, or you might deliberately induce it, but in general there is a problem of this nature, and I don't have any answer to it. I am only pointing to some fundamental questions here about what we mean.

I am not going to try to answer the question "Should we have the CVM?" I think you can see my attitude is very sympathetic; there are a lot of difficulties in CVM and there are a lot of difficulties in any kind of measurement which purports to do the same thing, for example to give values appropriate for welfare judgements. Also, in my few brushes with actual environmental analysis or health analysis, it appears to me that in the estimates produced by our technological colleagues -- our medical colleagues, our engineer friends -- errors on the order of one to ten are considered to be perfectly normal. On one such project on which I was

associated, for example, they were asked "What is the effect of nitrogen oxides emitted by supersonic transports on the ozone content of the stratosphere?" Well, the chemists had some laboratory experimental data, but they didn't know how long the nitrogen oxides would stay in the atmosphere. They didn't know whether the same chemical effects would occur because the reaction took place in the presence of a large mix of other chemical species that might upset the situation. There were some other factors involved. Although the effect they expected was there, there were other effects due to the supersonic transports that they hadn't allowed for. These scientists were perfectly aware of the limitations of their knowledge, and there were many more problems, problems which will turn up in that or any other effort.

The question is, should we be disturbed if we think that our error is within the factor of plus or minus fifty percent, or even double that? Let's talk about ratios of 3:1 or 5:1; compared to the other sources of ignorance in most of these environmental fields or the technological ignorance, and basic science ignorance, is this something to worry about, is this one of the biggest sources of uncertainty inside the environmental assessment?

C. COMMENTS BY PROFESSOR DANIEL KAHNEMAN

The "State of the Arts" document (Chapters I-VI) is an impressive piece of work. I was struck by the close correspondence that is sometimes observed between directly assessed market values and estimates derived from people's answers to hypothetical questions. Although psychologists commonly have greater faith in hypothetical questions than economists do, I was surprised that it was possible to do so well with the CVM method. I was also impressed with the intellectual rigor and honesty of the analysis. The critical task is to specify the conditions under which the CVM is likely to be valid and useful. Indeed the Reference Operating Conditions (ROC's) that are listed in Chapter VI define restrictions, warnings, or caveats on the use of this method. I would like to add a few more. It is my impression that several restrictions that were not mentioned in this volume should be considered. The purpose of my remarks is to suggest new ROC'S, to ensure that the use of CVM be constrained to problems in which its results can be trusted. To emphasize the continuity of my concerns with those of the authors of the book, I shall continue their enumeration of ROC's, in adding to the four that they stated.

1. Reference Operating Characteristic #5: The CVM should only be used for problems that have a "purchase structure." Let me now define what I mean by a "purchase structure." I distinguish two structures of transactions: purchase and compensation. In a purchase somebody pays to obtain one of two general kinds of things. People pay for improvements, gains, goods and services that make them better off than they were; they also pay to prevent a normal and expected deterioration. It is perfectly normal for a patient who has an illness and expects to get worse to pay for a treatment that will preserve her current level of health. I describe transactions of this general kind as having a purchase structure. Transactions that have a different structure often occur in the context of environmental affairs. In what I call a "compensation structure," we start with somebody who has an endowment -- for example a nice view, or clean air -- which is threatened by some deliberate and optional action of other people. Giving up this part or aspect of the endowment will make the individual worse off than before. The individual is requested, and sometimes coerced, to sell part of his or her endowment, in order to benefit someone else or society at large.

It is not always easy to determine whether a problem has a purchase structure or a compensation structure. The key diagnostic is whether the change in the individual's endowment is a normal, expected, and natural event, or an optional and therefore avoidable one, which only occurs because some economic agent or some social institution chose to follow a particular course of action. The optional and voluntary nature of the loss of endowment defines transactions that have a compensation structure. Let me illustrate the distinction by an example. Trees can be lost either to pests or to human action. Thus, a beautiful view may be ruined because a virus has attacked the trees, or because someone is logging or mining the area. What is the value of the view to the individual who is threatened by its loss? I wish to defend the controversial idea that the value of the view is not the same in these two situations. The loss of the view to the

pests, which the individual might pay to prevent, creates a purchase structure. The loss of the view to someone else's voluntary action naturally creates a compensation structure. If someone makes me worse off, I expect to be compensated.

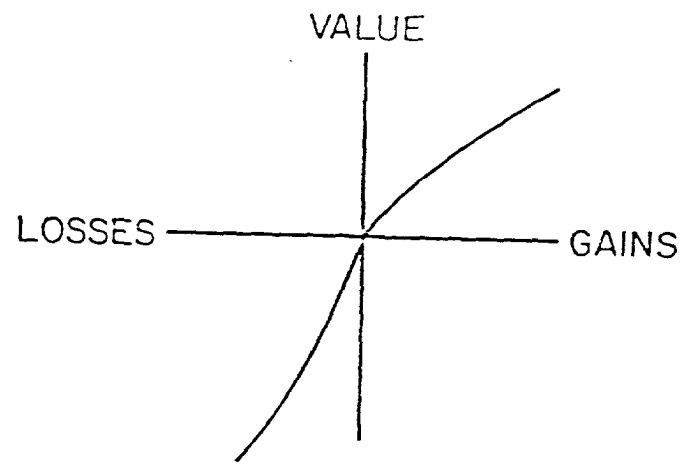
There is an obvious relation between the two structures of transactions that I have distinguished and the two methods of evaluation commonly used in CVM: willingness-to-pay (WTP) and willingness-to-accept-compensation (WTA). Standard economic theory assures is that the values assessed in the two ways should differ only by a (usually small) income effect. Because it is clear that the use of WTA measures in CVM often yields obviously absurd numbers, the spirit of previous chapters is to allow using WTP measures as a substitute for WTA measures, even when the transaction that is contemplated has a compensation structure for which WTA is appropriate. I have to make it clear at the outset that I do not favor the use of the WTA, which I believe to be very problematic. However, I suggest a restriction on the use of the measure that is favored by most of the authors represented in this document: "Willingness-to-pay should not be used as a measure of value in transactions that have a compensation structure." The proposed restriction is based on the idea that the value of the difference between two states depends on the cause of this difference, and on which of the two states is considered normal. Thus, the same loss of view will not have the same value if it is caused by a pest or by the intervention of a government agency. This is a psychological claim which, if accepted, has significant implications both for CVM and for public policy.

I shall try to defend this position, which may strike many of you as heretical, on the basis of theory rather than data. Specifically, I want to relate the idea to a central aspect of a theory of choice -- prospect theory -- that my colleague Amos Tversky and I have developed (Kahneman and Tversky, 1979, 1984; Tversky and Kahneman, 1981). The theory includes an analysis of value that compares each valued outcome or attribute to a neutral or normal reference point (See Figure 12.1).

Improvements or gains appear to the right of the reference point, and the value of all improvements is positive. Deteriorations and losses appear to the left, and their value is negative. The value function in the Figure is drawn crudely in two segments, with the function distinctly steeper in the domain of losses than in the domain of gains. The figure illustrates the phenomenon that we have called "loss aversion" (Kahneman and Tversky, 1984): losses generally loom much larger than corresponding gains.

To give you a sense of loss aversion, try comparing the intensity of the pain of losing \$50 to the pleasure of finding \$50. In another context consider a simple gamble, where on the toss of a coin you stand to win or lose a certain amount, with equal probability. The caution with which people approach such gambles far exceeds what could be explained by a concave utility function for money. For example, when I asked my students what minimum prize would induce them to put a \$10 stake on the toss of a coin, the average amount they demanded was over \$25. There is no way of deriving such extreme loss aversion from any sensible notion of utility for wealth, but the observations are readily explained by assuming that a gamble on even odds only becomes acceptable when the possible gain is inflated to compensate for the much higher sensitivity to possible losses -- as illustrated by the slope of the function of Figure 12.1 in the positive and in the negative domains.

Figure 12.1: A Hypothetical Value Function



To see the contribution of loss aversion to our story, consider the difference between dirty air and clean air. I argue that this difference can be legitimately evaluated in two different ways, depending on what is viewed as the normal reference point. First consider an individual who 'has dirty air. He lives in one of the areas of Los Angeles that are most afflicted by smog, and has now been offered the opportunity to purchase clean air, perhaps by moving somewhere else, or possibly by paying a share of a public clean-up project. In this case of a purchase structure, I propose that the difference between dirty air and clean air should be valued on the positive side of the value function.

Now consider an individual who lives in an area where the air is clean. Clean air is the normal state of affairs for this individual, but now a company wishes to move in, and to take action that will pollute the air. This case has a compensation structure. I propose that the same difference between dirty air and clean air should now be valued on the loss limb of the value function, which happens to be a great deal steeper. Thus, the value of the same difference between clean air and dirty air depends critically on where one is coming from. Note, however, that the present state of affairs does not always determine the relevant neutral reference point. For example, if the air is currently clean but is expected to get dirty from natural causes, as in the case of trees that still look good but are actually dying from a disease, the reference point is adjusted at least in part to the anticipated change. Gains and losses are probably relative to a state that is expected for the near future, rather than to the status quo.

If loss aversion is accepted as a fact of valuation, it follows that WTP is an acceptable method only for purchase transactions. In particular, WTP should not be used as a measure of value for people who are made to lose their clean air or their trees because of the intervention of some other agent. The fairest way to represent such cases is by recognizing that the experience is a genuine loss, and that the compensation should reflect this fact. I do not recommend using the WTA method to estimate this value, because I agree with the recommendations of the panel that this method is likely to produce useless results. My point is only that the use of WTP is likely to yield serious underestimates of the value of a good in a compensation structure.

There is a fair amount of evidence for the phenomenon of loss aversion on which the present argument rests. There have been many reports, in your own literature and in other contexts, of the so-called buying-selling discrepancy (Gregory, 1982; Kahneman and Tversky, 1984; Knetsch and Sinden, 1984; Thaler, 1981). This discrepancy can manifest itself by a difference between buying and selling prices, or by other measurements of reluctance to trade.

Among the examples of buying-selling discrepancy discussed in the present volume, the 3:1 ratio of estimates of WTA and WTP for hunting permits appears to be very solidly documented. It indicates, in the present terms, that the value of a hunting permit is not the same if one is receiving it or giving it up. Another striking example is that of the effects on housing values of formally designating some areas of California as high in the risk of earthquakes. When people who discovered that they lived in such a region were asked how much it would be worth to them not to face the risk -- that is, how much they were willing to pay to have the same

quality of life in an area that is free of that risk -- the value was about \$5000. When people in other areas were asked what sum might induce them to move to a designated high-risk one, the estimate was \$28,000. This huge discrepancy cannot be explained by self-selection. It is probably produced in part by some people who say "I won't do it, I would never willingly accept the risk!" The frequent refusal even to entertain the idea of a trade is one of the banes of the WTA method.

I repeat these examples in the present context to emphasize the idea that loss-aversion, the buying-selling discrepancy and reluctance to trade are highly robust effects that we ought to accept as such. It does not appear tenable to argue that, simply because economic theory says that there should be no difference between WTP and WTA, then there is no difference. This is one of those cases in which, when there is a conflict between observations and theory one should give the observations a chance.

The discrepancy between buying and selling is not a universal effect -- it can be made to vanish experimentally, and it frequently vanishes in the real world. What are the conditions under which we may expect no discrepancy between WTP and WTA? Reversible transactions offer one obvious example in which a large discrepancy simply makes no sense. The money that is spent to buy a loaf of bread is surely not evaluated as a loss. The 2:1 ratio for the values of losses and of gains, which is suggested by observations in just acceptable gambles, is certainly not applicable to routine payments. The attitude to the downside of transactions may change for recurrent reversible exchanges, in which one becomes familiar with the experience of getting a thing and giving it up. What is given up is eventually perceived as an opportunity cost rather than as a loss, and loss aversion is then not a factor.

When a loss is imposed on an individual on a unique occasion, however, there is no reason to expect the evaluation of gains and losses to be so balanced. Can we legislate that an individual is not allowed to have a steeper value for losses than for gains, at least in unique and nonreversible transactions? I submit that it is not reasonable to legislate preferences to that extent. We must therefore pay considerable attention to the buying-selling discrepancy when it exists. When it does, and when the problem has a compensation structure, the use of WTP to measure value must, in my opinion, be avoided. Tricky issues will arise, of course, because of the complex mixture of objective and subjective considerations in the problem. How should we evaluate trees that are taken out to permit mining, but were doomed anyway by a pest? Is the individual allowed to ignore the fact (if indeed there is such a fact) that utility bills may rise significantly unless the trees are torn down? Obviously, the determination of the neutral reference point cannot always be left to the individual, but the fact remains that there are situations of genuine and legitimate loss, for which a WTP measure will not provide a fair assessment.

Let me repeat in closing this topic that I have not spoken as an advocate of the WTA measure. Indeed, my aim was to raise a problem rather than offer a solution: by restricting the scope of CVM to measures of willingness-to-pay in problems that have a purchase structure, we may have restricted the application of the method quite substantially. There are surely many cases of compensation structure in which we would like to measure value, but the measure of WTA is suspect and WTP is not an acceptable substitute. The development of adequate methods of evaluation for

such problems is for the future -- and it will require much hard work.

2. ROC #6: The use of CVM should be restricted to user values, rather than to ideological values.

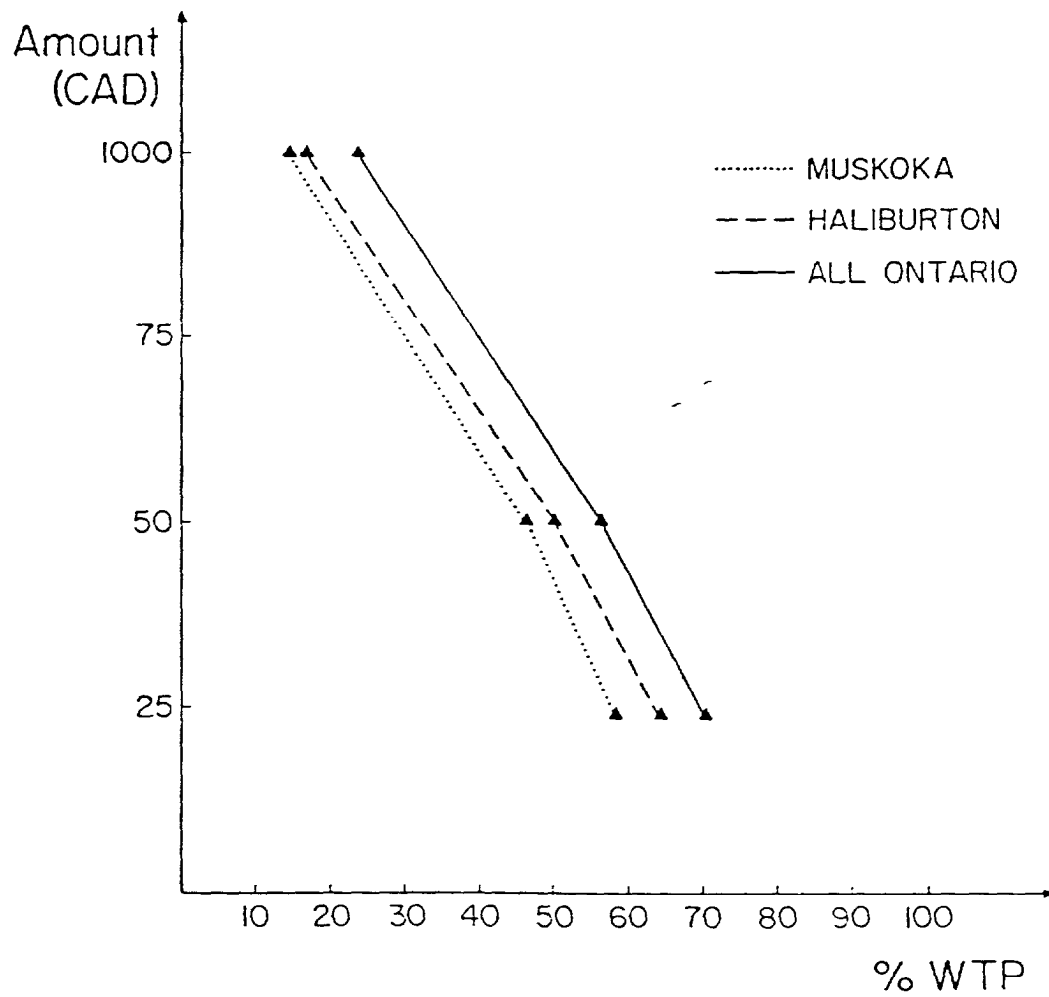
The thrust of this suggestion is that we should exercise great caution in measuring option values and reservation values, because the responses that are obtained in such measurements are likely to be heavily loaded with ideological content. To illustrate the notion of ideological loading, I shall quote from telephone surveys that Jack Knetsch and I have been conducting among the residents of Toronto, in which they were asked WTP and WTA questions about a number of hypothetical environmental changes. The key observation is that there is a class of problems in which people's answers to preference questions seem quite insensitive to the numbers that are mentioned in these questions. Indeed, people seem to be ready with an answer before the relevant numbers are specified. Professionals who are skilled in analyses of tradeoffs know that it is not possible to give a sensible answer to the question "What is more important, health or income?" without specifying how much health and how much income is at stake. Naive respondents have no such difficulties and they may be expected (this is a question we have not, in fact, asked) to state a clear preference for health over income. Similarly, I suppose that naive respondents will have a clear answer to the question: "What is more important to making people happy at work, the challenge of the job or the quality of the social life?" The willingness to choose among inadequately specified options suggests that the possibility of tradeoffs is neglected. Preferences of this kind appear to reflect a hierarchy of ideological values.

It is reasonable to assume that the CVM, which is offered as a substitute for the market, is not intended to measure ideological values -- but it may nonetheless be contaminated by such values. How can such contamination be detected? Common sense is a help, of course, but more formal diagnostics can also be applied. I will describe one, which I call "symbolic demand."

Consider the three demand curves of Figure 12.2. First, imagine that the dotted line represents the proportion of customers who are willing to pay different prices for 10 pounds of apples, and that the dashed line similarly represents the demand for a pair of shoes. What can we say about the demand for a package that combines the apples and the shoes? The answer depends on the distribution of demand for apples and shoes and on the possible covariation of the two goods in the demand of individuals. The figure illustrates a special case in which the demand for both goods is about equal. If in addition the goods are independent, as apples and shoes probably are, the vertical sum of the two separate demand curves provides a fair approximation to the demand for the package. In any event, the solid line can only represent demand for the package if the two goods are entirely redundant, so that either on its own is as good as the combination of both.

The three curves of Figure 12.2 do not in fact represent demand for apples and shoes. Instead, they represent answers of three groups of respondents in our telephone survey, who were asked about their willingness to pay an extra tax to maintain the fishing in some regions of Ontario. The leftmost curve represents the proportion of respondents who are willing to pay \$25, \$50 or \$100 or more for cleaning up the lakes in the Muskoka region. The next curve to the right displays the willingness-to-pay for a similar

Figure 12.2: Expressed Willingness to Pay Tax for Cleanup to Preserve Fishing in Muskoka, Haliburton, and All Ontario.



cleanup in the Haliburton region, and the rightmost curve describes the willingness-to-pay to clean up all the lakes in Ontario. The demand functions for the three cleanup operations are strikingly similar.

The results indicate that people seem to be willing to pay almost as much to clean up one region or any other, and almost as much for any one region as for all Ontario together. We know from other surveys that these responses do not reflect expectations of personal enjoyment from the cleanup, since Toronto residents are willing to pay substantial amounts to clean up the lakes of British Columbia! People seem to answer such questions as if they had been asked "What do you want to do about keeping fish in our lakes?" and "How important is the issue to you?" The dollar number merely expresses the strength of the feeling that is aroused by these questions. Because the questions all elicit symbolic expressions of the same attitude, there is not much difference between the numbers that are attached to a single region and to all of Ontario. I suspect that this pattern is hardly unique, and would expect similar failures of summation of demand for other value-laden "goods," such as human lives that could be saved by social action: the hypothesis is that willingness-to-pay to save lives will be largely independent of the number of lives that are to be saved. I call this "symbolic demand" because it is true of symbols that quantity is sometimes irrelevant: a small flag can be as good a symbol as a large one. The economically incoherent pattern of demand illustrated in Figure 12.2 can be a helpful diagnostic of evaluations that are dominated by ideological commitments.

The main point of these remarks is to question an assumption. As an outsider, both to economic analysis and to the use of CVM, it is natural for me to ask "What are the basic presuppositions of the work reported in the present volume?" One central cluster of presuppositions is that there exists a set of coherent preferences for goods, including non-market goods such as clean air and nice views; that these preferences would be revealed by a proper market; and that these preferences can be recovered by CVM if only the biases in CVM are eliminated. I find these to be very strong assumptions. In particular, I question the existence of a coherent preference order at the individual level, which is waiting to be revealed by market behavior. I am not sure that I have a "true" dollar value for the trees that I can see out of my window; that the market defines the perfect way of revealing the true dollar value of the trees; that the only problem of valuation is to discover that dollar value; and that it is therefore the task of methods such as CVM to achieve estimates of the market value.

An alternative way of looking at things would start from the assumption that preferences are often shaped by the eliciting procedure. This is, I think, the real significance of the starting point bias, about which so much has been said in this volume. For example, Jack Knetsch and I have tried a number of starting points in questions about the value of cleaning up lakes. We found that the proportion of respondents willing to have their taxes increased by \$50 to clean up the Ontario lakes varied from 18% to 64% depending on the starting point. The implication of this huge bias is that the respondents have no clear idea of how to answer the valuation question and that they consequently clutch at straws. One of the straws that is provided is the dollar amount that is mentioned in the question. Let me suggest a hypothetical reconstruction of the thinking that a respondent may do in answering a valuation question. "They ask whether I would be willing

to pay \$25 to clean up the lake. I have no idea, really, but \$25 is probably a number that divides the population about equally. What I do know about myself is that I seem to feel (more/less) strongly than many other people on environmental issues ... I feel the government isn't doing enough ... or there are too many environmentalist crazies blocking economic progress for the sake of fish and ducks." The initial Yes or No could well be determined in this manner and the magnitude of the anchoring bias suggests that it often is.

By the way, there is sad news for anyone who thinks that the bidding card will eliminate the problem. Several recent studies by Jack Knetsch and Robin Gregory have confirmed the highly predictable result that the bidding card is susceptible to anchoring biases. Responses obtained with a bidding card are unlikely to be free of anchoring biases, for the simple reason that the range of values on the card provides information. Indeed, the middle region of the card is a hint about what the experimenter considers a reasonable answer to the questions. There is no magic way of preventing respondents from latching onto such weak hints as they may find in a question, when they have no better way of answering it.

A specific recommendation about CVM use may be in order here. No study of CVM should be conducted without manipulation of the potential anchors or suggestive numbers in the valuation question. Furthermore, these manipulations should be powerful enough to elicit the anchoring effect in all its beauty; it is all too easy to fail to find a significant bias by using a biasing manipulation that is too weak. The use of the anchoring results depends, I suggest, on the magnitude of the bias that is observed. If the bias is small or moderate, values obtained with different anchors can be averaged to obtain an improved estimate. If the bias is large, however, a different conclusion may be in order: When the estimates are too susceptible to anchoring or to starting point bias, perhaps we should stop our analysis right there. Like the incoherent pattern of demand that was discussed earlier, extreme susceptibility to suggestive numbers may be taken as an indication that the dollar values that we hope to measure simply do not exist.

Doubts about the existence of a coherent preference order are not only raised by anchoring biases, and are not restricted to non-market goods. Tversky and I have studied a wide variety of choice problems in which preferences are highly susceptible to what we call framing effects: preferences are affected by inconsequential variations in the descriptions of options (Kahneman and Tversky, 1984; Tversky and Kahneman, 1981). Framing effects violate a principle of invariance, which Kenneth Arrow has called "extensionality." Framing effects are probably common in studies of CVM. Any demonstration that preferences are susceptible to such effects in a particular context would raise doubts about the applicability of the method to that context.

In the early days of CVM, one of the main concerns was with the possibility that respondents may wish to disguise their true values, for strategic reasons. A more realistic concern, I submit, is that users of CVM often deal with people who simply do not have the kind of coherent preference order that the theory assumes -- especially in domains for which they lack market experience. The cautious recommendation is to avoid using the method in such cases.

3. ROC #7: Accurate description of payment mode is essential to the CVM.

My final point echoes a remark that Kenneth Arrow made earlier, to the effect that preferences are highly sensitive to procedures as well as to outcomes. This, as Ken has pointed out, is perfectly rational. It may not have been emphasized to a sufficient degree in the treatment of CVW in the present volume. The social arrangement within which the payments in WTP are going to take place is an essential aspect of the payment method, and I put that as my last ROC.

The classic theory of public goods incorporates an idealization that one should not forget. The theory adopts the assumption that I urged you earlier to reject: that people have a specifiable demand for the good in question, and that the task in public-good demand estimation is merely to aggregate the demand of all the members of the community. The aggregate demand or the aggregate WTP is then accepted as the value of that particular public good. If you are beginning to be suspicious about this assumption, then some qualifications are in order. In particular, it is likely that the value of a particular product of social action to an individual depends strongly on the details of how that action is performed -- for example on the equity of the distribution of payments.

There is a bind here: we intend the CVM to mimic what a free market would generate. But a free market is inconceivable for many of the goods that we wish to value. The only realistic way to achieve some goods is by government intervention or by social action, and the cost of this action must be distributed, either progressively or equally, among members of the community. In such cases, it is indeed impossible to separate the value of a good from the procedure by which that good is obtained. In particular, WTP will then depend on others' payments. Note that this is a concern for equity, which is not the same as a strategic attempt by individuals to minimize their payments and maximize their benefits. What happens here is simply that if I am asked to pay \$50 to preserve Ontario fish, I would like to know who else is going to pay \$50. This is a legitimate concern for a person to have, but it is one that severely constrains the validity of the CVM: the value that is estimated when a particular social arrangement is assumed by the respondents may not be transferred to another.

In conclusion, there are cases in which the CVM in effect provides a market survey for a good that could indeed be marketed -- the more successful applications of the CVM appear to be of this kind. However, when we deal with goods that can only be provided by the public, the survey, whether we like it or not, actually provides an estimate of the results of a referendum on a special-purpose tax, or on the fair allocation of a particular good. This view of the CVM has implications that extend even to the proper statistical analysis of survey results. My impression is that the tradition of using the mean of WTP derives from the idea that the quantity to be estimated is the total demand for the public good. Total demand is naturally assessed by estimating average individual demand, which is then multiplied by population size. If what we have is actually a pattern of voting on a policy question, then the median amount that people are willing to pay might be just the measure that we want. My suggestion is not that the median should always be used. The point that I wish to make is that the statistics that we employ must be adapted to the structure of the decision problem, and to the structure of the social mechanism by which the public good will be provided.

D. COMMENTS BY PROFESSOR SHERWIN ROSEN

The study is a very useful one that lays out the picture very clearly and completely. Speaking as someone who has a small stake in some of these issues and whose a priori views tend towards skepticism, the report made a convincing and positive case for the CVM.

Three little criticisms refer to some "cheap shots" that detract from the document as a whole in my opinion. One concerns a quote of Joan Robinson's in Chapter I, to the effect that there is no possibility for empirical truth in economics. That may or may not be true, but what is the virtue of raising it in this context? Besides, the quote was just naive in terms of empirical controversies in other sciences.

The second point concerns the discussion of social welfare measurement (Chapter II), where a suggestion is made that market prices don't reflect values. I fail to see the point of unqualified statements of this sort. The authors are all economists and they should take the thorough economic point of view. Let other experts take different positions. Distrust of the market often appears in environmental protection discussions and is popular in some quarters. But the proper audience to influence first is economists, and economists won't take this position. Apart from externalities there are cases where market prices don't reflect social values involving taxes and other distortions, neither of which are mentioned and could be taken into account.

The third point concerns raising very general questions about the validity of utility theory and rationality. Again I don't see any payoff for that in this context because I don't see what alternative there is to utility theory in a cost-benefit calculation, and cost-benefit theory is all we have to go by in this business. Besides, there are tests of rationality in this context, e.g., integrability tests.

Now, on to the main points. There is little question, as I said at the beginning, that the CVM approach is a promising one and a progressing research program. Sometimes there is a flavor in the report of some Olympian battle among methods here. Yet the question is extremely well posed: How much are people willing to pay for certain things? What we are trying to achieve is a good method of answering that question; the question to be answered isn't controversial at all. These methods are not really mutually exclusive. I certainly don't see then that way. We shouldn't be looking for the Best Method; a universal Best Method probably doesn't exist. One method shouldn't be excluded over the other, because the best empirical research looks at the problem in alternative ways and through varieties of evidence. The more varieties of evidence we have, the more assured we will be of the correct answer. Another value of this enterprise is the value of learning how to do survey research. Economists have little skill at survey research though we certainly use much survey data generated by people in other professions which is not necessarily ideal for our purposes. I am very hopeful that some of the work here will spill over into other aspects of survey techniques in economics.

We particularly need more evidence on validity and reliability of the method. In this respect I found Chapter VI of the report the most interesting. It is the only one that gives really hard numbers on a comparison of this approach with some others that leads to some indication of validity or reliability. On this, I think the authors sometimes use

difficult theoretical arguments when the numbers speak for themselves, and no theorizing is needed.

I would like to suggest an additional approach: The use of replication studies. I don't see any evidence where a contingent market had been replicated. Such studies may be boring, but if we are doing experimental work here of this sort, I think you have to get some replication. I would like to see how the "goose study" done in Oregon compares with the one in Wisconsin, and perhaps in some other place. These repeated trials are an important way of learning how valid the method is.

I also would like to make a point on this WTP and WTA difference, since I strongly disagree with Kahneman on the interpretation of Brookshire's study on earthquakes. To my mind there is a basic confusion here between whether preferences are inconsistent -- whether indifference curves exist and so on -- and whether there are differences in preferences among subjects. People who live on the fault will answer a question differently than people who don't live on the fault. This is how I read the description of the Brookshire study. People who don't live on the fault are more worried about earthquakes and require much larger compensation to live there than the people who choose to live there. They have different preferences, and if one is labeled WTP and the other labeled WTA, you are heading into big trouble. There is a study by Glen Blomquist about the value of lake views in Chicago, where someone who lives in the high rises right on the lakeshore was asked "How much would it take to get you to move off the lakeshore?" How much would they have to be paid to give up their lake view? The response was a lot different than the amount that people who didn't have a lake view would be willing to pay to get a lake view. It is obvious that the people who didn't have a lake view self-selected themselves -- they didn't care that much about it.

Another point that deserves emphasis relates to the strategic hypothetical bias argument. The point attributed to Rick Freeman in the volume is important and bears repeating. There is no strategic bias so long as the CVM is strictly hypothetical. If it is hypothetical, then the respondent knows his answer won't affect any policy, and there is no incentive to misrepresent preferences. But if it is hypothetical, there is no great incentive to go through the effort and cost of sharp calculation to elicit true preferences. This is the real conundrum in the method and underlies my initial skepticism about the CVM. It is worrisome that there are only four or five studies where one can make empirically meaningful comparisons. Now, one can argue theoretical points until doomsday, but we need some more empirical comparisons to check the validity of these methods. In this respect also, I don't view the hypothetical bias argument as so ill-defined as the authors suggest. It is an economic argument, a cost benefit question on the cost of calculation in answering a question. It seems difficult to test this. The authors want to make a formal test of the proposition; but I don't see how a true test can be devised except by comparison with some alternative method.

I found the section on accuracy (Chapter VI) to be unclear. Perhaps I missed something, but the 50% number that was derived for assessing accuracy appears ad hoc. Precisely what scientific argument was used to arrive at that number? The 50% figure also seems to imply that people don't know their own minds. Suppose that we had a perfect CVM, as good as we could make it, and a person could calculate down to the last nickel how

much a project is worth to him. Why isn't that a fairly accurate number? Why should it be valid only up to 50%?

I also would argue with some of the supporting textual material concerning this point. The results on the variation in estimates of demand elasticities, discussed in Chapter VI, are not all that interesting, since not all of those studies are equally valid. For example, everybody's estimate of the demand for sugar or whatever should not be counted in calculating standard errors. Some of those studies are awful and should be thrown out of court. They are no good. Some are much better than others.

Let me give you an example. In standard demand theory some years ago, a well-known study rejected the theory of demand because the Slutsky matrix wasn't negative semi-definite, on translog specifications. People have reworked that very same data -- it was aggregate time series data -- using much weaker revealed preference tests rather than a translog system. Revealed preference tests never reject the theory of demand. There is not enough price variation to get true revealed preference comparisons in the actual data and all the budget sets are nested. So what apparently happened in that study is that the translog analyses imposed a lot of curvature on the data that just wasn't there. That curvature was invalidly imposed as a maintained hypothesis, and it came out wrong.

Let me close with some questions that I don't feel were addressed by the study, that perhaps should be. One concerns the scope and limitation of the method. What kind of problems is the method best addressed to and used for? Where would we be most comfortable in using it? Goose hunting is one thing, but how about nuclear hazards, nuclear power radiation, promotion in the Southwest for fossil fuel generation and so on? Not only do we need clarification on where these methods might be more useful; but also whether they should be confined only to environmental issues. Perhaps they would be useful for other kinds of public goods decisions, the size of the military for example.

Another question that wasn't addressed is the cost of implementing the method relative to alternative methods. Perhaps other methods are cheaper. We need more information on this. Surveys are expensive, and we are not told how expensive these surveys are.

The third point has to do with 'selectivity effects'. The earthquake site case is one example of it. The on-site experiments on CVM certainly select users by their taste. Let me go back to the goose hunters -- I was thinking while that was described that I would be willing to pay a few bucks to prohibit all goose hunting. I don't want to get shot when I go to view the Canada geese. More seriously, what is the relevant population for a survey in this area of research?" How does this relate to such things as protest votes, refusniks and so on, and precisely what is their role in the method?

The fourth point concerns the question of strategic bias which might arise if this technique was put on line and seriously used on a large scale. While reading the report, I had a vision of everybody hooked up via their PC's, direct on-line with EPA in Washington, making Groves-Ledyard votes one hour per day every day. If this technique gets serious and widespread use, we might well expect the results on strategic bias and so on that we are getting from current results to be invalid. At least I'd worry about extrapolation.

The fifth point is that the report, perhaps, adopts a fairly naive

approach to economic policy. In fact, it is the approach I would have taken myself four or five years ago, before I'd been exposed to the work of some of my colleagues, especially Stigler and Becker. We really have to address the political economy of EPA and other kinds of regulations. This is the kind of regulation that seems to use very little economic input. There are uniform standards, very little price incentive, and a lot of other things that apparently can be rationalized only by political considerations in pressure group politics. This raises questions of how the respondents act when they answer these questions. Do they take these kinds of political considerations into account? Is that another potential form of hypothetical bias?

E. COMMENTS BY PROFESSOR VERNON SMITH

Kerry said that one of the disadvantages of going last is that everything has already been said. But that is not really true for an experimentalist who goes last, because we nearly always have some data that we can show. I do want to show some data a little later on that are taken from experiments based on joint work with Peter Knez and Arlington Williams. These have to do with the subject of calibration. We are studying private goods market situations, but we are also asking WTP and WTA questions.

As economists, our primary tool for solving a problem is to think about it. This leads us to slip, perhaps unconsciously, into the assumption that economic agents also solve their decision problems by thinking about them. In testing decision-theoretic propositions by interrogation methods, I think psychologists and others seem also to have assumed that the economist models the decision maker as a consciously analytical agent. This seems to be implicit in procedures that ask subjects to choose among a set of alternatives. Yet, I think the typical subject in a market experiment, based at least upon my experience, does not appear to operate in this manner. For example, some subjects "learn" over time to adopt demand-revealing dominant strategies, but they really couldn't articulate why they do this. Some never learn; some seem to latch on to it right away, but I think they would have a lot of difficulty explaining to you why.

In more complicated experimental markets than the simple auction, subjects really learn to do quite well for themselves, and also for the theory of competitive markets, without having an understanding or even a perception of the market as a whole, which is anything like our rigorous models of market analysis. This strongly suggests the possibility that rational behavior may not be consciously calculating. Specifically, it suggests the hypothesis that direct decision responses from individuals based upon thinking about alternatives may lead to violations of the principles of rational behavior, but what individuals actually do in the sequential replicating market context may not violate those principles. Hence, people may in some sense learn to be rational through market experience.

Now, in Chapter VI we find a report of some laboratory experiments by Coursey, Hovis and Schulze, which show clearly that what people say about WTP and WTA is not necessarily what they do asymptotically in a repetitive market experience. I want to emphasize the importance of this hypothesis and these corroborating results for any program that will apply to the contingent valuation method, by briefly discussing some similar experimental results that involve a rather different market context than those used by Coursey et al. Let me begin by providing some reinterpretation of WTP and WTA data as it applies to estimating the value of a particular good, such as the right to avoid tasting sucrose acetate, which is, I think, the commodity used in the Coursey et al., experiment. Or the right to hunt a goose or a deer. In discussing the difference between WTP and WTA measures, I think it is important to distinguish between differences for the same individual and differences among individuals, and I have a feeling that has been confused in the discussion. I think the former has been claimed to violate rational choice theory if there is a "large" difference between WTP and WTA, though "large" is not

very well defined as I read this literature. Most of the observers seem to find that such differences are larger than they expected. But this assessment is really subjective. For example, Coursey, Hovis and Schulze note that the income effect should be small since WTP and WTA are small relative to income. Well, I think Don Coursey should remember the subject at the University of Arizona who, when she collected the \$25 to \$30 earned in a market experiment, commented that I had just saved her a pint of blood. Now, people who derive income from blood sales seem unlikely to satisfy the assumption that income effects will be negligible.

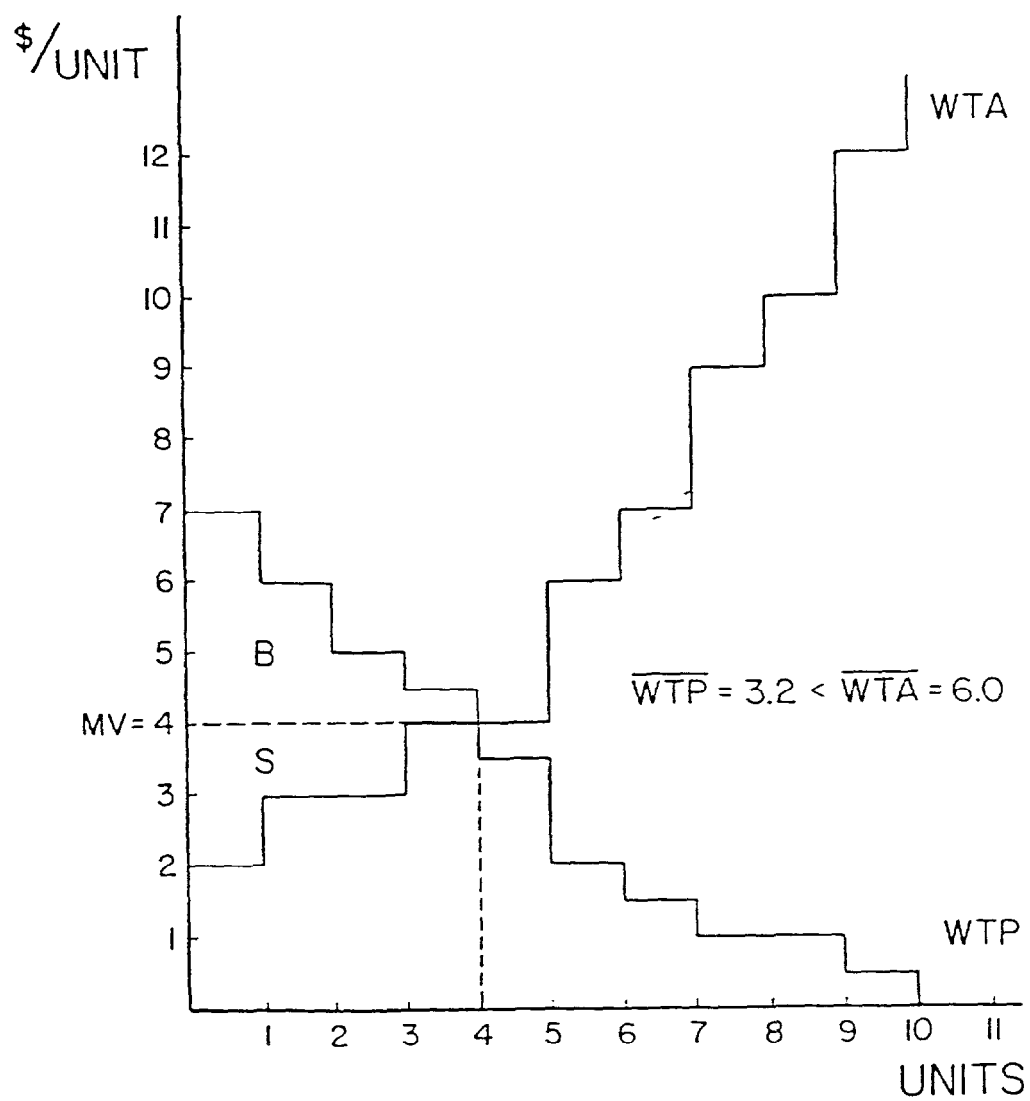
I think it is well to bear in mind that all these speculations here are just highlighting the fact that we really don't know. The guy says the divergence is larger than he expected based on the theory, but that requires an interpretation that might be incorrect. Now, differences in WTP and WTA across different individuals, even if large, should not disturb us, since that is the kind of divergence in valuation that is the basis for exchange. Large differences may simply mean that we can expect to observe low volume in market trading. The point here is that unless the distributions of WTP and WTA are disjoint, across individuals, there will be no gains from trade.

Insofar as the CVM is used to value private, non-traded goods, such as goose hunting and deer hunting permits, it seems to me that the objective is to measure market value, which can be quite different than mean WTP or mean WTA. Let me illustrate what I have in mind. I give you the standard freshman diagram, which I am going to use to lead into some of the experiments that I am going to report. In Figure 12.3, the downward sloping line is a set of WTP measures that you might get by interrogation from a group of individuals and the upward sloping line is a set of WTA measures that you might get from the same group and it shouldn't surprise anyone that the mean WTP might be different than the mean WTA; or that both of these might be different from the market value (MV) -- the value that maximizes the gain from exchange. In Figure 12.3, area B is buyer's surplus, area S is seller's surplus, and $B + S$ is the total surplus from competitive market exchange.

The experiments I am going to tell you about were not set up as WTP or WTA experiments. They had a quite different purpose; in fact, the study had been going on for six or seven months before it occurred to me that it might be a good vehicle for asking WTP and WTA questions. The experiments involved studying rational expectations theory in an asset trading context. In these experiments, twelve subjects might participate in an asset trading market, and each subject is given an endowment in cash and an endowment in securities. One subject might get \$9.50 in cash and no securities, another might get \$5.00 in cash and one security, and so on. The securities all pay a random dividend with a distribution which is known to everyone. The understanding is that after each period of trading we will draw from this dividend distribution and everyone who holds some inventory of securities will receive that dividend, with everyone receiving the same dividend.

We have been using this vehicle for looking at rational expectations theory, as I mentioned, and we also wanted to use it to see if we could create market bubbles and crashes in the laboratory. Parenthetically, I'll mention that we began with an assumption that it might be very hard to do this. It turned out that we were quite wrong, it was very easy to do in these finite games -- in a fifteen period game we had people's expectations

Figure 12.3: WTA - WTP Relationships



of capital gain causing them to bid up prices in a bubble-like market, sometimes followed by crashes from the price peak as some began to wonder if they would be able to find another "fool" who would pay the high prices they had just paid. This, at least, is our interpretation of the results.

It occurred to us that a simple version of this game might be a good environment in which to look at WTP and WTA measures. Suppose everyone has gone through the instructions in one of these experiments and each knows their initial up-front cash and securities endowment. Then we can ask them the maximum they would be willing to pay for an additional unit of

securities added to their inventory position; or what is the minimum they would accept to sell out of inventory. Our thought was to ask them these questions -- hypothetical questions -- and then put them in a single period of trade and see what trades occur and observe the actual trading prices. Now maybe we will get off-the-wall answers to the WTP and WTA questions, but on the other hand the resulting hypothetical market value might not be a bad predictor of trading prices. If we got the results, for instance, shown in Figure 12.3, in a particular survey, those results would predict, on the basis of an interrogated supply and demand, that the mean price in the market will be around \$4.

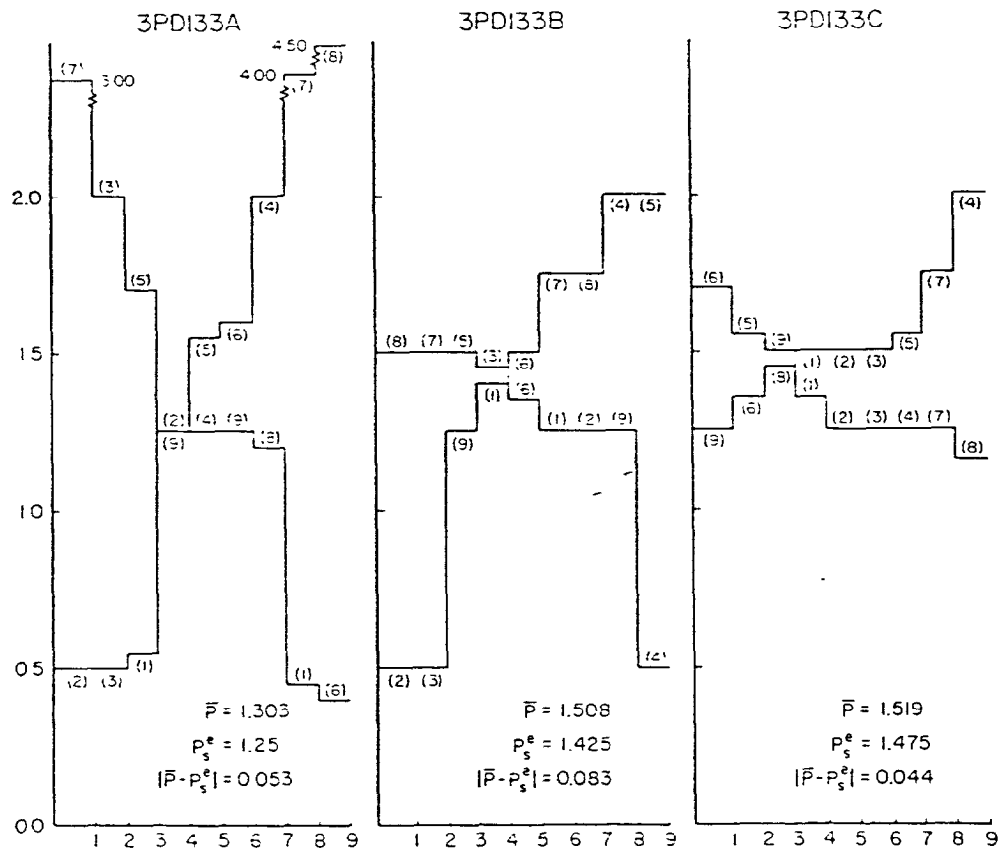
So that was one of our questions: How good a predictor is this hypothetical vehicle, even though there might be a lot of evidence of some sort of irrationality in the answers to these questions -- the point being that it is possible that our theory of preferences is bad but that our theory of markets is not so bad. That is, our markets may do a pretty efficient job, given whatever preferences are, even if those preferences do not conform to our a priori expectations based on expected utility theory, or what have you. Another question, and this one relates to the Coursey et al. study, was that we wanted to see whether, if there were some wild choices in WTP and WTA responses, these would tend to disappear, and get more reasonable, as the subjects obtained market experience.

Figure 12.4 shows you some responses to hypothetical WTP and WTA questions that we asked nine individuals who are about to trade a simple gamble. The questions were put, and answered, prior to observing these people trade. The gamble has an expected value of about \$1.25, paying \$.50 with probability 1/2 and \$2.00 with probability 1/2. We got some "crazy" answers here -- referring to Panel A, Figure 12.4, someone says they're willing to pay \$3.00 for this gamble! For Subject 7, the WTP was \$3.00 and the WTA was \$4.00. Subject 2 will sell for \$.50 -- that is, WTA was \$.50 -- but was willing to pay \$7.25 for an additional unit. You can see that some responses are all over the place. In fact, the mean willingness to pay is \$1.39, the mean willingness to accept payment is \$1.83, and the predicted price is \$1.25, the expected value of the gamble! There is an old principle in economics that the cutting edge of the market is what the marginal sellers and buyers are going to do. It doesn't make any difference if YOU have some wild intramarginal WTP answers as long as they are balanced by comparable WTA answers. You may have these kinds of responses, and yet the market as a whole may not be making such an irrational prediction as to what's going to happen. Here, in fact, the prediction of these interrogations is the same as what rational expectations predictions would be -- namely a price of \$1.25.

After these questions were asked, the subjects traded. They followed New York Stock Exchange trading rules: any buyer can make a bid, any

Figure 12.4

Tentative Results from University of Arizona Experiments



\bar{P} = mean observed price

P_s^e = predicted price based on stated WTP and WTA

seller can make an offer, for a single unit. If either a bid or an offer is accepted, acceptance becomes a contract. The subjects make the market -- there is no auctioneer, except in the form of a rule. There are various kinds of rules governing the market, and the participants must subscribe to them, but there is no conscious intervention by any kind of super-agent; the subjects are doing all the trading. In trade, the mean price on that market was \$1.30, compared with the predicted price (by both interrogated supply and demand and rational expectations), which was \$7.25.

At the end of the first period of trading we reinitialize everybody with the same endowments of cash and securities that they had before, and we ask them the same questions again, and Panel B of figure 12.4 shows the answers they gave us. WTP and WTA are starting to tighten up, but they are predicting a higher prize. The market clearing price on the basis of the hypothetical interrogations is now about \$1.42. As it turned out, that wasn't too bad a predictor of what they did, since the mean price we observed in trading was \$1.50 -- quite a bit above the expected value of the gamble. In fact, both the prediction by the WTP and WTA measures and the actual market were well above the expected value of the gamble. Most of these experiments were repeated five times. In this particular case I will just show you results for three periods.

Panel C of Figure 12.4 gives results for the third period. The interrogation (hypothetical) procedure predicted about \$1.48, and the mean we observed was \$1.52, again both above the predictions of the rational expectations model.

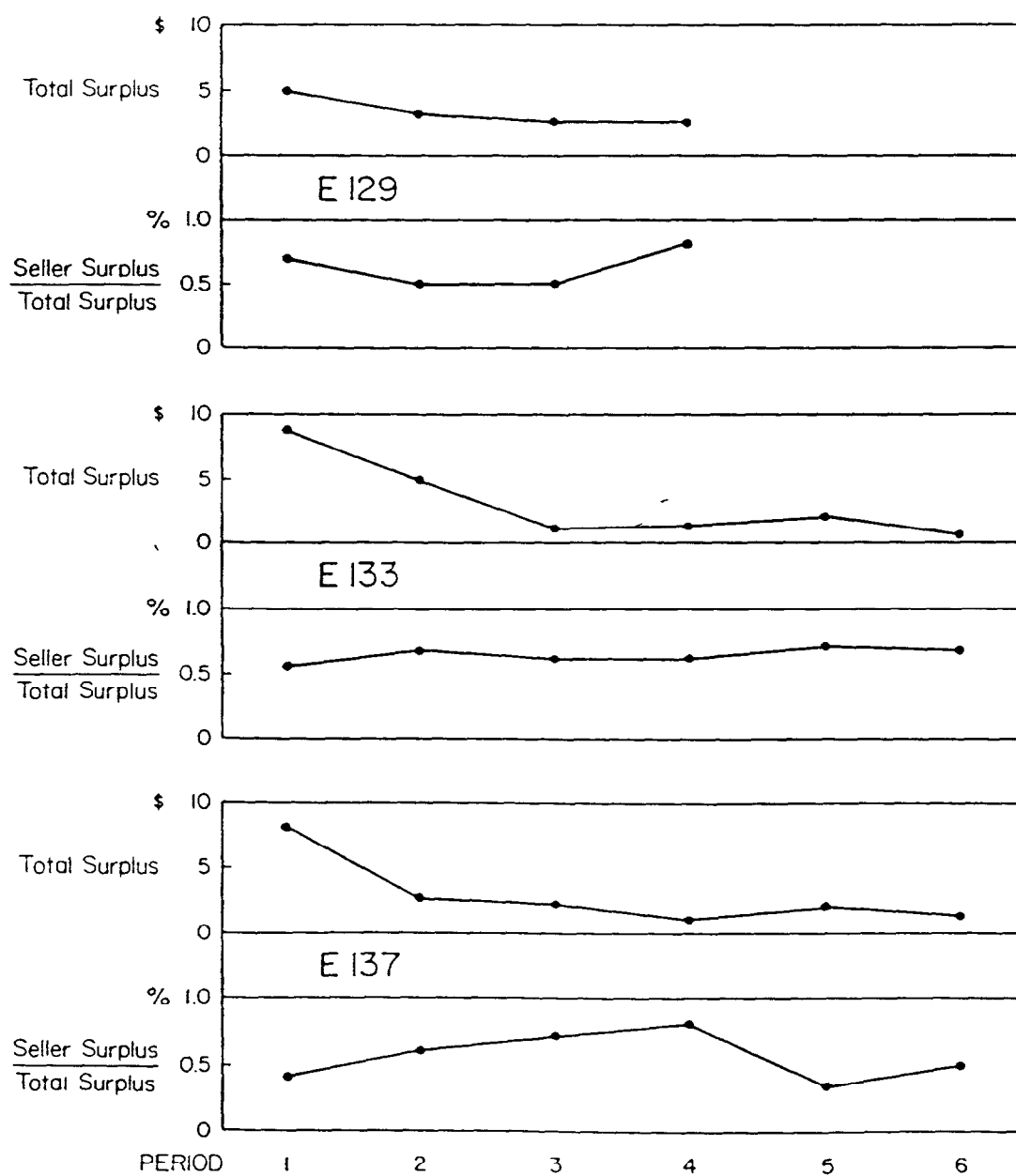
In Figure 12.5, if you look at how total surplus changes, it seems to me you see something of how much people and preferences are coming together across three trials. You can see that total surplus is falling. Most of the decline is over by the second or third iteration. After this it stabilizes.

Hypothetical WTP and WTA are certainly not an accurate predictor of what the people do. They do poorly in predicting volume -- each interrogation provides a prediction of what the volume of trade will be, and volume was nearly always higher than that. But across all replications, the hypothetical WTP/WTA measure does better than the rational expectations prediction as to what the mean observed price will be.

Now we ask the question that Coursey et al. asked, to see whether in our case, as in theirs, most of the adjustment came from the WTA side, with the WTP remaining quite stable: How does the seller surplus change relative to total surplus? Referring to Figure 12.5, you can see that we do not have evidence (in terms of the surplus measure) that most of the adjustment was coming from the seller side. Actually, we haven't computed the means of WTA and WTP yet. We hadn't seen those means as particularly significant, because we were thinking in terms of private goods, of course, but we'll do that and maybe the means are adjusting more on the WTA side than on the WTP side.

Let me close by coming back to a point made by Ralph d'Arge. I think it was said that the real test is whether economists can come up with proposals for introducing markets in the allocation of environmental goods. As I read the CVM work, it seems to me that what you are mainly working on is proposals for some sort of a substitution for the market, a calculation substitution. I really think we ought to devote a little time to thinking about whether there might be the possibility of creating markets where they

Figure 12.5: Change in Seller Surplus Relative to Total Surplus



don't now exist, and let the market do the calculation. In fact I have a proposal, one which involves the estimated 45,000 wild horses and asses that are in Nevada, Arizona and Utah. These have been very controversial -- apparently the population is growing at the rate of about 12% a year, and of course there is heavy pressure on grasslands -- the ranchers want to shoot them all and Wild Horse Annie wants to save them all. There is an 'Adopt a Donkey' program, which I understand works pretty well for the attractive ones, but the ones that are ugly, well, they just can't find anyone to adopt them.

So I have a very simple proposal: We take the membership of the Sierra Club and the Friends of the Earth and other environmental organizations and also all the members of the Cattlemen's Associations of Utah, Nevada and Arizona, and distribute among them at random 45,000 options to shoot a horse. Then we list these options on the Pacific Stock Exchange, and allow them to be traded in an open market. And so, if a rancher wants to shoot a horse, he has to buy one of these certificates and then shoot the horse and turn the certificate in. An environmentalist who wants to save a horse buys one of these certificates and sits on it. So the idea is you use the market to manage that stock of feral animals. To keep the floating stock of certificates equal to the stock of animals, the U.S. Wildlife Service would estimate the animal stock, say every five years. If the animal stock exceeded 45,000 less the number of redeemed certificates, they would just declare a certificate dividend such that the certificate stock equals the number of feral animals. If the animal stock was less than this figure, they would declare a negative dividend in certificates to maintain the equality. This would allow the stock of certificate claims to keep pace with the net biological change in the animal stock. Sherwin Rosen says he would be willing to pay something to keep anyone from shooting the geese -- if he feels the same about these animals, all he needs to do is go out and buy all 45,000 of the certificates.

XIII. SUMMARY AND CONCLUSIONS

A. OVERVIEW

The reader has at this point been exposed to our initial assessments of the CVM (Chapters I - VI) as well as to reactions regarding those assessments by a wide range of scholars, along with their assessments of the CVM, presented at the Assessment Conference and reviewed in Chapters VII - XII. We were particularly impressed with the quality of the interchange between Conference participants concerning new and provocative ways of looking at where we are and where we might go in terms of the development of the CVM. In addition to the invited responses to Part I of this book which are given in Chapters VIII - XII, we received comments on the pre-Conference draft of Part I by Drs. R.C. Mitchell and R.T. Carson (Resources for the Future, Inc.). The points raised by Mitchell and Carson in their Comments represent substantive contributions to the assessments of interest in this book; thus, with Mitchell and Carson's permission, their comments are included in an Appendix to this chapter. As will become quickly apparent to the reader, the insightfulness of Mitchell and Carson's comments is reflected by our repeated references to them throughout the balance of this Chapter.

Our task now is to draw together our discussions of the CVM in Part I with those by Conference participants in Part II to the end of suggesting final conclusions as to the state of the arts of the CVM as a method for valuing public goods. As an aside, the reader should understand that in response to the many constructive criticisms of our assessments of the CVM offered by Conference participants, we have not altered the pre-conference conclusion set out in Part I -- we have left the "warts" in our earlier discussions and conclusions as they were. Thus, as we develop final conclusions in this chapter, the reader is part of -- can participate in -- the intellectual assessment process wherein constructive, critical interchange between scholars is used to mold conclusions which ultimately reflect (we hope) some degree of consensus. State of the arts conclusions regarding the CVM are developed below in the following manner. In Section B, attention is focused on the weight of structural bias in the CVM: biases which have been argued to result from such things as starting points, choice of payment vehicles, strategic behavior and information. Section C considers an issue that, in Chapter V, was presented as being of central importance for assessments of the CVM: the potential for hypothetical bias in CVM measures of value. In Section D we address the question of primary importance for our state of the arts assessment of the CVM: how does one evaluate the accuracy of CVM measures? These discussions are brought together in Section E wherein we consider the bottom line: What is the state of the arts for the CVM? The chapter concludes with Section F wherein we define critical issues for future research with the CVM.

B. STRUCTURAL BIASES IN THE CVM

Four structural types of bias in value measures derived with the CVM were given particular attention in earlier chapters of this book and were of particular concern at the Assessment Conference. These potential biases, discussed in turn below, are: strategic bias, starting point bias, information bias and vehicle bias.

1. Strategic Bias. In general, the views of Conference participants concerning strategic bias in CVM measures parallel those developed above in Chapter V. Freeman notes the absence of strong empirical evidence for free-riding behavior, which in his view suggests that individuals will not behave strategically in purely hypothetical or contingent market settings -- a point of view seconded by Rosen. Professor Arrow finds neither theoretical arguments nor empirical evidence compelling in terms of strategic behavior by CVM subjects.

Both Freeman and Rosen emphasize, however, the potential dependence of the "no strategic bias" conclusion on the fact that, within hypothetical settings, subjects in the CVM study are not offered obvious opportunities to manipulate outcomes; i.e., as noted in Chapter V, the potential for strategic bias is less, the more hypothetical the valuation process in the CVM. Such dependence, if it exists, raises two related problems, however. First, and most obviously, a trade-off is suggested between strategic bias and hypothetical biases -- this issue will be discussed in detail below in Section C. Secondly, a number of researchers are currently advocating alternative structures for the CVM wherein emphasis is placed on the subject's perception that his/her response will influence policy. Thus, Randall's theoretical model (Chapter VIII) is based on the assumption that subjects believe that the results of the valuation exercise will influence policy; within this framework, the "penalty" for a non-preference-researched response is argued to be that the subjects' opportunity to influence policy is wasted or misused. Such focus on influencing policy, as noted by Randall, is suggestive of referendum formats; indeed, Kahneman views the CVM as it stands as effectively simulating a referendum. Carson and Mitchell (Appendix) look to referendum formats -- political markets -- as an alternative framework for the CVM and as a means for identifying "reference operating conditions" relevant for assessing the accuracy of CVM measures (Appendix, part 4).

Ceteris paribus, the use of referendum-type formats as a means to investigate hypothetical bias may be questioned on the grounds that the more real is one's perception of the relevance of his/her responses in terms of influencing policy, the greater is the potential for strategic bias (see, in Chapter XII, Rosen's "personal computer" analogy). It is not clear that such is the case, however. As implied by Carson and Mitchell, couching the CVM within the context of a referendum may in fact amount to the adaptation of the CVM to an institution which differs markedly from the market institution which common applications of the CVM attempt to simulate. The possibility of tying the CVM to alternative institutions (vis-a-vis the market institution) is an interesting and potentially important point and is considered in some detail below in Section F.

2. Starting Point Bias. In Chapter III we noted that when the CVM valuation process is initiated by the interviewers' question: "Would you be willing to pay \$X," post-bidding valuations tended to cluster around \$X.

The dependence of CVM values on the initial or "starting point" value of \$X was described as a "starting point bias." We noted empirical evidence supporting the existence of such biases -- Carson and Mitchell (Appendix, section 2.a) suggest still stronger evidence for such biases and argue that studies suggesting the absence of such biases may be flawed by the low power of tests used to examine hypotheses concerning starting point bias. At least two methods have been suggested for eliminating/mitigating starting point bias: the use of a payment card (c.f. Chapter III), and Freeman's naval gunfire analogy of "bracket and halving" (Chapter X).

Professor Kahneman (Chapter XII) proposes quite a different context for treating and interpreting starting point bias. Kahneman suggests that the finding of starting point bias is indicative of a CVM "commodity" for which subjects are unable to answer valuation questions. For some types of commodities, lack of experience or familiarity with the commodity results in subjects' having great difficulty in putting dollar values on the commodity -- subjects are not "hiding" anything from the interviewer nor are they attempting to be clever, they simply do not know how to answer the valuation question in a meaningful way. Thus, rather than adopting means to eliminate starting point biases, Kahneman seemingly views means to identify the existence of such biases as an important part of the study design: the presence of such biases indicates that subjects are too ignorant of the commodity to be able to value it meaningfully, in which case the CVM should not be applied to the commodity in question. Kahneman offers further "sad news" (XII. C): use of a payment card does not eliminate the problem inasmuch as value ranges on the bidding card provide the potential for "entering biases" (indications of "reasonable" responses).

When starting points are used in CVM studies, we concur with Carson and Mitchell that the evidence suggesting starting point biases is indeed compelling. While, as is discussed in Section C, Kahnemans' concern that a subjects' lack of experience/familiarity with a particular environmental good may result in his/her having difficulty in placing monetary values on the good -- indeed, "familiarity, and/or experience is an ROC in Chapter VI -- received empirical evidence does not seem to support the notion that such difficulties are made manifest by starting point biases. Following Mitchell and Carson's suggestions (Appendix), higher powered tests for such biases may well result in starting point biases showing up in CVM studies involving commodities with which subjects are reasonably familiar -- see the seven studies wherein derived CVM values are shown to compare favorably with values derived from indirect market methods (Table 6.12). Thus, we would argue that starting point bias may well reflect other phenomena, e.g., the subjects' interpretation of starting points as indicative of actual costs for a proposed environmental improvement. Moreover, it would appear to us that payment cards can be structured so as to eliminate the potential for the "entering biases" of concern to Professor Kahneman. Thus, while an issue of concern, the authors conclude that starting point problems should be amenable to control through care in the design of the CVM payment card.

3. Information Bias. In Chapters III and V, the authors pointed to the confusion that one finds in the literature as to the substance of what is referred to as "information bias;" at the heart of this confusion is the failure on the part of many writers to distinguish between effects on CVM valuations arising from the subject's exposure to more information ("more" in quantitative and/or qualitative terms) regarding the commodity or

valuation process as opposed to the subjects exposure to different information -- "different" in the sense that two sets of information imply two different market (valuation) structures or two different commodities.

Randall (Chapter VIII) suggests that such confusion is eliminated as follows. Rational subjects base their contingent market decision on (i) the value of the commodity offered; (ii) the rule by which the agency decides to provide or not to provide the commodity; and (iii) the rule that determines the payment to be exacted from the subject. Since, according to Randall, only (i) is relevant for valuing nonrival goods, the pertinent question is: do (ii) and (iii) encourage accurate reporting of (i)? In this vein, Randall argues that different information which affects (ii) or (iii) should affect reported measures of willingness to pay. Such changes in information then result in effects on WTP measures that are expected a priori. Such effects, therefore, are not biases. In this manner, Randall rejects the notion of "information bias."

Related to Randall's point (iii) -- as well as to (ii) -- is the design question as to whether or not a subject in the CVM should be given information concerning bids by other subjects. Arrow argues that such information should not be given due to the potential effect of this information in eliciting strategic behavior. Moreover, Arrow views such "second hand" information as possibly leading to biases resulting from subjects' dependence on more informed judgments of others, as implied by their bids. Freeman argues that such information could lead, in effect, to a form of starting point bias. Along a slightly different line, Kahneman sees information concerning (iii) as an integral part of the valuation process -- any one individuals' "true" willingness to pay is inextricably related to what all other individuals are paying for the commodity in question, i.e., Kahneman implicitly rejects the economists' commonly-used assumption of independent utility functions.

However, Randall's arguments concerning (i) - (iii) address only one part of the sources of information of concern in Chapter V: changes in information affecting value structures and/or commodities; his arguments do not seem to speak directly to the relationship between reported valuations and the quantity/quality of descriptive information concerning the commodity. In these regards, it would seem that in cases where systematic differences in valuations are associated with changes in the quantity or quality of information describing the CVM commodity, the implied "bias" may well be attributable to difficulties in "information processing" described in Chapter V. Arrow points to the difficulties in balancing the potential benefits of providing subjects with descriptive information with the subject's difficulties in processing that information. Freeman sees such biases as positive vis-a-vis assessments of the CVM inasmuch as they may be interpreted as indicative of subjects' approaching the valuation process in a meaningful way; i.e., subjects use information provided to form perceptions of the CVM commodity and base their valuation responses on that information.

Thus, in terms of information which has the effect of altering the nature of the CVM commodity, rules for providing the commodity and/or rules which determine actual payment, we would concur with Randall's judgment that one would expect such changes to alter bids, in which case a bias per se is not implied. On the related subject concerning a subject's exposure to bids offered by other subjects, we find the argument that such information

may result in undesirable biases compelling; in this regard, we note that, while a substantive issue which perhaps warrants future inquiry, Kahneman's rejection of the assumption of independent utility functions weakens results from virtually all benefit assessment methods. Finally, in terms of biases which may result from different levels of purely descriptive information given to CVM subjects, two concluding observations appear salient. First, an integral part of pre-tests of questionnaires must be the effort to balance the subject's need for information with his/her general capacity to absorb -- process -- the information. Secondly, as suggested by Freeman, one must avoid interpretative generalizations of CVM results to environmental changes other than those specifically described in the CVM instrument.

4. Vehicle Bias. Conference participants, particularly Professors Arrow, Kahneman and Randall, took sharp issue with Chapter V's discussion of vehicle bias. The essence of our discussions of vehicle bias in Chapter V is reflected in Freeman's (Chapter X) statement of the vehicle bias problem: our inability to determine which payment vehicle, if any, provides "true" (unbiased) values and which payment vehicles lead to biased values. Arrow, Kahneman and Randall argue that the search for an unbiased payment vehicle is misguided -- "biases" are not implied by systematic variations in offered values and payment vehicles.

The essence of Arrow and Kahneman's argument (see Kahneman's ROC Number Seven in Chapter XII.C) is that the social arrangements by which payments are to be made -- the payment vehicle -- is an integral part of the CVM commodity per se, i.e., one cannot separate the value of the commodity from the procedures by which the commodity is provided and payment is made. Of course, this is Randall's argument (iii) concerning information bias which was discussed above. In this regard, Kahneman rejects the notion that values based on one set of "social arrangements" may be transferred to a different set; Arrow sees differing preferences -- and therefore values -- related to purchases via use permits, general taxation and/or general price effects, as rational. Thus, Arrow suggests that WTP depends on the structure of "P".

These arguments are surely compelling and have important implications for the design of and interpretation of results from the CVM. First, following Kahneman (Chapter XII.C), reflecting the fact that our commodity is not a market commodity, but a commodity which can only result from social action (government intervention), the CVM'S mode of payment is selected on the basis of realism -- what payment vehicle would most likely be employed, in fact, if the commodity were to be provided? Secondly, paralleling Freeman's interpretative limitations related to information bias, we explicitly acknowledge, without apology, the potential dependence of obtained valuations on the adopted payment vehicle.

5. Conclusions. In terms of the potential structural biases in CVM values which this Section addressed, the current state of the arts in the CVM may be described as follows. First, all else equal, strategic bias does not appear to be a major problem in applications of the method. Two caveats are relevant for this conclusion, however. Interactive information concerning other subjects' values, as might attend efforts to bring standard CVM practices together with experimental techniques, may introduce incentives for strategic behavior. Further, efforts to reduce the potential for hypothetical bias (discussed below) in the CVM, a la

Randall's proposed dependence on a subject's belief that his/her response will actually affect public policy, may invite strategic behavior in applications of the CVM which rely on market institutions -- the implications of structuring the CVM in alternative institutions are discussed below in Section F.

Secondly, the authors submit that the use of carefully structured payment cards can effectively mitigate starting point bias in applications of the CVM involving commodities with which subjects have had some degree of market-related experience -- where subjects are reasonably "familiar" with the commodity. For other commodities, Kahneman's concern with starting point bias -- with or without a payment card -- may be well-founded, but it is unclear to the authors how one would distinguish between anchoring-sorts of biases in these cases and biases attributable to the myriad hypothetical-related issues concerning decision-making under uncertainty, attitude/behavior and others which arise when individuals begin at the bottom of a learning curve relevant to an environmental commodity.

Thirdly, the "information bias" rubric seems to serve no useful purpose for assessments of the CVM; indeed, it may be counterproductive. In terms of the quantity/quality of descriptive information concerning the CVM commodity, it seems reasonable to expect that pre-tests of questionnaires can be used to balance information needs with information processing capacities for "appropriate" commodities. Once again, the familiarity issue arises as does the relevance of the authors' suggested ROC's. In the case of unfamiliar goods, in the authors' minds, it appears sanguine to expect that processing capacities can be balanced with the bulk of information that might be required to elicit reasonably informed valuations from subjects.

Finally, in terms of information concerning rules pertaining to the provision of the commodity and/or to payment, we see little to distinguish these information "biases" from those considered under the rubric of "vehicle bias." In these regards, we consider the state of the arts as one wherein the notion of vehicle bias, broadly defined, is without substance. One acknowledges that such rules are an integral part of the valuation process. Values derived via the CVM are then interpreted as simply applying to the specific commodity described in the questionnaire, provided under the "social arrangement" (rules for provision and payment vehicle) described in the questionnaire. In this context, one views with equanimity the rational fact that different payment/provision institutions -- social arrangements -- may result in different valuations.

C. HYPOTHETICAL BIASES IN THE CVM

The reader will recall the many "faces" of hypothetical bias discussed in Chapter V. As one might expect after reading that chapter, the issues associated with hypothetical bias, and the implications of such biases, served as a source of interesting exchanges at the Assessment Conference. Reflecting some degree of consensus among conference participants, the major issues related to hypothetical bias, as they are relevant to our state of the arts assessment of the CVM, are: the preference research issue(s); the comparability of WTA and WTP measures; and the attitude v. intended behavior issue. Those issues are considered in the discussions that follow.

1. Preference Research Issues. Under the rubric of "preference research" developed in Chapter V, three distinct lines of argument can be discerned from the Conference papers and discussions: the role of incentives for accurate valuations; the importance of a subject's familiarity/experience with the CVM commodity; and the (related) learning issue.

(a) Incentives and accurate valuations. In V.B above, arguments by Freeman (1979) and by Feenburg and Mills (1981) concerning the lack of incentives for "accurate" valuation responses in the CVM were distilled into a hypothesis of the form: valuations with actual payment equal valuations without actual payment (i.e. with hypothetical payment). Underlying this hypothesis was Freeman's notion that, since individuals suffer no utility loss from inaccurate responses to CVM valuation questions, they lack incentives to engage in the mental effort (and consumption of time) required to research preferences and formulate meaningful evaluations. Our review and interpretation of the literature related to the above hypothesis -- primarily the works by Bohm (1972), Bishop and Heberlein (1979), Coursey et al. (1983) and Slovic (1969) -- resulted in our conclusion that results from research to date belie the above stated hypothesis, i.e., substantive differences in values result when real and hypothetical payments are involved. Obviously, the implications of this conclusion would not bode well for the CVM. If hypothetical payment does not provide incentives for accurate responses in the CVM, and absent means for quantifying such biases, the viability of the method may be seriously questioned.

Mitchell and Carson (Appendix) take sharp issue with our conclusion. Based on their reworking of data used by Bohm and by Bishop and Heberlein, they find that results from these works concerning actual/hypothetical payment are much weaker than those reported in the authors' original papers. In turn, however, we should note Bishop and Heberlein's critiques of Mitchell and Carson's reworking of their data, given above in Chapter IX. Moreover, Mitchell and Carson challenge the relevance of results from the Coursey et al. study inasmuch as the study's focus is on WTP-WTA differences, and results related to actual/hypothetical payment differences are simply inferential. Finally, referring to the literature in cognitive psychology, their discussions with Slovic suggest that, first the general literature on this topic shows equivocal findings; and second, that results from Slovic's 1969 study do not strongly support the sweeping conclusion offered by us in Chapter V.

Of course, Mitchell and Carson do not argue that hypothetical payment does not result in bias; rather they argue that the question remains open. Arrow seemingly agrees that the question is open. He argues (Chapter

XII-B) that in the pseudo-reality of the CVM, well-structured questionnaires which create real-like markets may well be capable of generating real-like results. Randall (Chapter VIII) offers a stronger argument: notwithstanding hypothetical payment, incentives for a subject to research preferences and formulate accurate valuation responses are provided by the subjects' concern with foregoing an opportunity to influence policy -- we have noted above the potential conflict between this position of Randall's and the strategic bias issue noted by Arrow, Freeman and Rosen. Perhaps still stronger in these regards are results from laboratory experiments conducted at the University of Arizona reported by Vernon Smith (Chapter XII-E). Based on these experiments, Smith concludes that interrogated WTP/WTa values (corresponding to hypothetical payment/compensation) were found to be better predictors of post-trading equilibrium values for prices than a priori predictions from expected utility theory. Moreover, while pre-trade predictions of trading volumes were typically inaccurate, Smith notes that predicted (hypothetical) valuations were generally close (around 95%) to actual market-clearing prices.

There remain, however, the results of Bishop and Heberlein's recently completed study of Sandhill deer hunting permits (Chapter IX). As in their early goose-hunting permit study, Bishop and Heberlein find significant differences between bids involving cash and hypothetical payments in all of their WTA experiments (Table 9.2) and in three of the four auction formats used in their WTP experiments (Table 9.3). Based on these findings, Bishop and Heberlein conclude that the evidence for bias related to hypothetical payment is rather convincing. Moreover, they argue, no matter how closely the Reference Operating Conditions are met, hypothetical bias (attributable to hypothetical payment) will remain.

Bishop and Heberlein's conclusions, as well as the results from their impressive Sandhill study, are not readily dismissed. No matter how weakened by Mitchell and Carson's analysis, there exist research results from several studies (reviewed in Chapter V) supportive of those offered by Bishop and Heberlein. But there exists a great deal of evidence which challenges the weight of Bishop and Heberlein's conclusions. In this regard, we note the above-cited observations by Mitchell and Carson and by Arrow, as well as, particularly, the experimental results reported by Vernon Smith. Moreover, results from Chapter VI's analyses of seventeen comparison studies demonstrates remarkable (in our view) consonance between values derived with the CVM and values derived from indirect market methods -- a degree of consonance which is, at worst, inconsistent with the full weight of Bishop and Heberlein's conclusions, particularly as their conclusions refer to commodities which to some extent satisfy our ROC's. Similarly, these demonstrations argue against the strong conclusion suggested by us in Chapter V.

In offering, then, a state of the arts conclusion concerning the incentives issue generally, and biases attributable to hypothetical payment particularly, the authors feel compelled to soften their conclusions in Chapter V and to concur in principle with Mitchell and Carson: at worst, evidence from research to date provides equivocal results concerning the hypothetical payment issue; at best, for public goods which satisfy the ROC's, evidence from comparative and experimental studies suggests that minimal biases in CVM measures may result from hypothetical payment.

(b). Familiarity/experience as a prerequisite for CVM commodities.

A second preference research issue developed in Chapter V concerns the extent to which subjects in the CVM interview can place meaningful Values On commodities with which they are unfamiliar -- they have no experience in trading/valuing the commodity in question. Hypotheses related to this issue developed by the authors in Chapter V focused on time and information requirements by subjects if they were to research preferences in a meaningful way to the end of formulating accurate valuation responses. In our search for research results relating to these hypotheses, myriad problems associated with such things as cognitive dissonance, mental accounts, information processing -- more generally, bounded rationality -- we were compelled to conclude that results from the received literature offered little that would support the notion that subjects, during the relatively brief period of the CVM interview, could define their preferences for a new, unfamiliar commodity in any meaningful way -- thus, our use Of ROC's 1 and 2 developed in Chapter VI.

The familiarity issue, and our requirement for experience/familiarity with CVM commodities as a Reference Operating Condition, was the subject of considerable controversy at the Assessment Conference. Freeman (Chapter X) essentially accepted the familiarity/experience issue as being on equal footing with the hypothetical payment/incentive issue as a potential source of bias in CVM measures, and expanded the familiarity argument in the following way. In contrast to conventional theory, Freeman argues that individuals have more accurate knowledge of their preference orderings in the neighborhood of those consumption bundles that they have actually experienced. In instances where individuals are moved into unfamiliar regions of their preference orderings, accurate preference orderings -- and therefore accurate valuations -- will result only after the individual can learn (via trial and error experiences) about this "new" region of consumption bundles. Thus, if the CVM involves small changes around neighborhoods of experienced consumption bundles (the individual is, therefore, somewhat familiar with the commodity), valuation responses will be more accurate than for CVM studies involving changes (or new commodities) which move individuals to regions of preference orderings with which the subject has no experience.

V. Kerry Smith acknowledges the potential importance of the familiarity issue, but takes the argument along two somewhat different lines. First he argues that the relevant state of the arts is one wherein we can say little, qualitatively or quantitatively, about the implications of the familiarity problem inasmuch as we have no model of how individuals behave/respond in the CVM milieu; he notes Hoehn and Randall's (1984) interesting beginning in this regard, to which we would add the logic suggested by Freeman (Chapter X). Secondly, and somewhat curiously, Smith argues that, in accepting the ROC's which require that subjects be familiar with the CVM commodity and its (at least) indirect market exchange, we require that the subject's choice experience is the equivalent of his/her knowledge of the features (outcomes) of the implicit market; i.e., such CVM studies elicit the subjects' perception/estimation of implied market outcomes for hypothetical changes rather than the subject's personal valuation of the commodity.

V. Kerry Smith's latter point warrants a closer look. If the CVM commodity was a loaf of bread, the subject's knowledge of market outcomes (the price that bread commands in the supermarket) would surely be reflected in the subject's bid. But the familiarity requirement for public goods is

not this strong, nor is the requirement for indirect market experience. In Chapter VI's example of air quality in Los Angeles, satisfaction of the familiarity ROC was argued on the grounds that subjects were (i) aware of (familiar with) air quality differences in various areas in the basin, and (ii) that equivalent houses in areas with better air qualities would cost "more." Individuals may have rough ideas of how much more beach-side homes cost than the housing counterpart in Pasadena, but it would be heroic to assume their access to hedonic measures which attribute values to the myriad attributes of the beach-side house (proximity to beach, crime rates, etc., and air quality). Faced with the question: "Living in Pasadena, what would you pay for (beach -side) levels of air quality?", a basis for the subject's calculation of a market solution a la Smith is not readily apparent. Thus, while Smith's call for modeling efforts concerning individual behavior within the setting of the CVM is (and was, at the Conference) well-received, his assertion that CVM applications for commodities satisfying the familiarity ROC's imply the generation of implicit market outcomes, rather than an individual's revelation of preferences, is not (to the authors' minds) convincing.

Kahneman argues that the requirement of familiarity does not go far enough in terms of imposing limits on applications of the CVM which may lead to a priori expectations of reasonably accurate responses. In Chapter VI, the authors, in describing the implications of the ROC's, noted that the ROC's precluded the derivation of value estimates for unfamiliar, and uncertain, commodities, such as those related to option, preservation and bequeathment values. Kahneman suggests the use of a distinct ROC which precludes the application of the CVM for deriving any value with ideological content -- i.e., only user values should be the subject of CVM applications. In support of his argument, Kahneman draws on the notion of "Symbolic (or incoherent) demand." Symbolic demand reflects an individual's hierarchy of values which, Kahneman argues, must inject itself into any economic or political context. Manifestations of symbolic demand -- manifestations of ideological "loading" -- are seen in subjects' inability to differentiate between values attributable to related, but nonsubstitute goods; e.g., a subjects' inability to differentiate, in value terms, between improved air quality in area A, areas A and B, and air quality throughout the U.S. (this particular example of symbolic demand is found in Schulze et al. 1984, Chapter I). Thus, to the extent that familiarity and uncertainty ROC's do not eliminate all possible applications of the CVM to commodities with ideological content, we are asked to expand the ROC's to preclude such applications.

(c) The learning issue. While inextricably related to the familiarity question discussed above, questions concerning "learning" are sufficiently distinct to warrant their separate treatment. At issue in these regards is the efficacy of various methods and techniques in assisting subjects in the CVM to first, more effectively research their preferences; and/or secondly, to more completely understand the nature of the contingent market and incentive-compatible behavior appropriate for that market. Methods/techniques of concern in these regards are: the iterative bidding process; the use of repetitive valuation trials; and more generally, the transferability of techniques used in laboratory experiments to applications of the CVM.

A recurring theme through Chapters III - VI is the authors' view that

the iterative bidding process must be used in CVM applications if meaningful measures of subjects' maximum willingness to pay are to be derived. This admittedly strong view was based primarily on three arguments developed in those chapters. First, the heuristic argument (Chapters III and IV) that, at the outset, subjects may not fully appreciate the "all or nothing" character of the contingent market and that the bidding process "prods" the individual to more completely research his/her preferences vis-a-vis the contingent commodity; as in any auction, demands on the subject's judgment as to the extent to which he/she really wants the commodity, increase as the stated price increases. Secondly, results from experimental economics demonstrate that subjects require time and repetitive valuation trials before they begin to fully appreciate the nature and implications of the valuation process. Third, and finally, the considerable empirical evidence which demonstrates significant differences between initial, one-shot values and final values derived with the bidding process.

While acknowledging that initial, one-shot, bids may underestimate a subject's maximum willingness to pay, Mitchell and Carson (Appendix) reject the notion that the iterative bidding process solves the problem; in so doing, they challenge each of the three arguments used by us in developing our contrary conclusion. The heuristic "prodding" argument is turned 180 degrees to suggest that the bidding procedure may in fact "bully" subjects into bidding more, given their awkward social position of having to say "no" to the interviewer's inferred request for a higher bid. While agreeing that CVM scenarios should include iterative elements which permit learning, Mitchell and Carson argue that the iterative trials of experimental economics are unnecessary to accomplish this end, and moreover, do not make the case for using the iterative bidding process. The necessary use of iterative trials in experimental economics, they argue, may well be related to the nonintuitive, second-price auction institution. In terms of one's understanding of the WTP format, they point to the data presented in Table 4.1 of Chapter IV which shows (for WTP trials) minor differences in bids across the repetitive trials. Finally, the interpretative weight of our empirical evidence demonstrating differences between initial and post-bidding values is implicitly challenged by Mitchell and Carson by the question: "To what does one attribute the observed differences: downward bias (as we argue) or a "bullying" effect?

Bishop and Heberlein (Chapter IX) also criticize the "categorical conclusion" regarding the need for iterative bidding suggested by us in earlier chapters. Like Mitchell and Carson, they point to the weak statistical tests in demonstrations of bid differences with and without iterative bidding processes and report results of their analysis of three bidding game studies wherein starting and iterated bids are positively correlated with hypothetical payment, but not correlated with actual cash payments. Referring to results from their Sandhill study, Bishop and Heberlein suggest that iterative bidding encourages subjects to exaggerate their willingness to pay; one should note, however, that only one iteration was used in their study. Finally, noting that iterative bidding precludes the use of mail surveys in application of the CVM, they suggest as an "ultimate conclusion" that the iterative bidding process may simply not be worth the trouble and expense.

In Chapters IV and VI, the authors devoted considerable attention to developments in experimental economics and the potential promise of

laboratory methods/techniques used by experimental economists for structuring and testing questionnaires to be used in CVM field interviews; particular stress is given to the use of "Vickery Auctions" and tatonnement processes -- basic methods used in experimental economics -- as means by which more accurate responses might be obtained with the CVM.

Our enthusiasm for lessons learned from experimental economics, vis-a-vis their meaningful transferability to the CVM, was not totally shared by Conference participants. Bishop and Heberlein criticized our stress on the need to conduct laboratory experiments while ignoring the contributions of field experiments -- a position supported by Arrow. In chiding the authors' "one-sided" emphasis on the virtues of laboratory experiments they point to the highly simplified and artificial settings of all laboratory experiments, and question the transferability of such results to real-world situations -- a criticism echoed by Mitchell and Carson as well as by V. Kerry Smith.

The emphasis given to Vickery auctions and the tatonnement process in Chapter IV was found particularly disconcerting by a number of Conference participants. In terms of the Vickery auction -- a "discovery" viewed by Bishop and Heberlein as a red herring -- Mitchell and Carson (Appendix) as well as Bishop and Heberlein (Chapter IX) acknowledge the effectiveness of the method in assessing institutional structures for private goods involving actual exchanges (see also, V.K. Smith, Chapter XI, Section 4.C), but fail to see how the method is to be used for hypothetical markets for public goods wherein exchange is impossible; in this regard, these authors argue that our reliance on the Coursey et al. (1983) experiment, involving the private good SOA, does not support our general conclusions. Given the nonintuitive format of the Vickery auction, and (as we report in Chapter IV) the repetitive trials required for subjects to learn incentive-compatible behavior implied by the format, both Bishop-Heberlein and Mitchell-Carson question how such repetitive trials are to be implemented within the CVM framework (see, also, Freemans' remarks in Chapter X). Iterative bidding, these authors maintain, does not substitute for the repetitive exchange trials of the Vickery auction format. Similarly, in terms of our suggested use of tatonnement processes as a part of the CVM, Bishop-Heberlein assert that, for hypothetical public goods of interest for the CVM, Groves-Ledyard procedures for implementing such processes may not cause respondents to reveal true preferences and may result simply in increased costs, increased confusion and lower response rates. In this regard, reliance on tatonnement processes for the large groups of subjects generally included in CVM studies "boggles" the minds of Mitchell and Carson.

While we accept the "Red Herring" comment of Bishop and Heberlein in the spirit of intellectual mischief in which it was intended, we do feel that the role of experimental economics in contingent valuation research has been misunderstood, most likely due to a failure in our exposition in Chapter IV. Rather than serving as guidance for the structure of hypothetical survey questions for the CVM, the demand revealing mechanisms developed by public choice theorists and experimental economists show how to obtain value estimates which are close to "true values" in laboratory situations. It turns out that even in the laboratory, it is fairly difficult to obtain "true" demand revealing values. First, one must use an incentive structure such as a Vickery auction for private goods. However, this not sufficient. In addition, individuals must be given a number of repetitive learning trials to understand the auction mechanism and learn that demand revelation is their best strategy. Only by using both, a demand revealing mechanism and by allowing sufficient learning experience to accrue via repetitive trials, do about 70% of the subjects actually reveal demand in laboratory settings. Thus, based on their observations, the Bishop and Heberlein study (described in Chapter IX) which actually attempted to repurchase hunting permits likely did not reveal demand for hunting permits since no opportunity for repetitive learning trials was given to participants and subjects most certainly had no prior experience selling their hunting permits. It then follows that experimental economics sheds little light on Bishop and Heberlein's hypothetical values, but suggests their "true value" obtained from actual behavior may have been biased for reasons other than those acknowledged by them. The primary lesson from experimental economics is, therefore, concern methods by which values may be obtained which are demand revealing as a basis of comparison for alternative, hypothetical measures of value.

These discussions conclude our capsulization of the controversies surrounding the preference research issues: issues concerning the need for incentives for accurate valuations, the subjects' need for familiarity/experience with CVM commodities, and the efficacy of iterative bidding and methods/techniques drawn from experimental economics for assisting subjects in their preference research processes. As to the implications of these discussions for the state of the arts of the CVM, conclusions in this regard are but deferred until we have considered other issues related to hypothetical bias. Thus, the authors' conclusion concerning issues related to preference research are given below in sub-section C.4.

2. The Comparability of WTP and WTA Measures. In Chapter VI, the authors submit as a Reference Operating Condition for assessing the accuracy of CVM values, the requirement: "WTP, not WTA, measures are elicited." (VI.E). The rationale for the authors' imposition of this ROC was based on two related lines of argument. In Chapter III (Section 4) we note that in spite of theoretical arguments (which relate to private goods) that WTA should equal WTP, empirical studies (Table 3.2) consistently demonstrate wide divergences between WTA and WTP measures; generally, estimated WTA measures are orders of magnitude greater than estimated WTP measures (Table 3.2). In Chapter IV (Section C), we argue that such observed disparities between WTA and WTP may be attributed to cognitive dissonance, which in the context of IV.C's discussions, is reflected (via the Davis, et al. experiment) by subjects' failure to recognize dominant strategies in a

Vickery auction, i.e., in some cases, iterative trials, whereby subjects learn that full demand revelation is their dominant strategy, results in the convergence of WTA to WTP measures. Such convergence was found to generally obtain (in the Coursey et al. experiment) under nonhypothetical circumstances, but not under hypothetical circumstances, an anomaly attributable to the lack of a market-like environment in the hypothetical experiments. In retrospect, we note the implications of this finding for earlier-discussed criticisms of our enthusiasm for the use of Vickery auctions in the hypothetical setting of the CVM (Section C-1-c). We also note the consistency of laboratory results with Randall, et al.'s (1983) argument (also, see Randall's arguments in Chapter VIII) that WTP underestimates "true" values while WTA overestimates such values.

A considerable amount of interesting and constructive criticism Of Our WTA/WTP arguments and conclusions was offered by Conference participants. First, various participants questioned our attribution of WTA-WTP differences in hypothetical settings to "cognitive dissonance" and our implied reliance on results from iterative trials in one experiment (the Coursey et al. (1984) experiment) as a means for eliminating cognitive dissonance. Thus, Bishop and Heberlein question the lack of symmetry Of learning effects from iterative trials on WTP and WTA measures in the Coursey et al. experiment: iterative trials affect WTA measures but, seemingly, not the WTP measures. Moreover, Freeman (Chapter X) questions our attribution of WTA-WTP differences to "cognitive dissonance" and the link between cognitive dissonance and our learning-via-iterative-trials arguments. In this regard, cognitive dissonance refers to the beliefs of a subject (on which preferences are based) which are persistent over time and in the face of contrary "facts," and which are changed by subjects via their selection of information sources which are consistent with "desired" beliefs (Akerlof and Dickens, 1982, p. 307). Thus, all else equal, the cognitive dissonance argument would lead us to expect little if any changes in bids with additional information (learning; Arrow, 1982). In these terms, a subject's lack of understanding of a Vickery auction (or any other valuation institution) may be viewed as distinct from an individual's value -related beliefs which are subject to cognitive dissonance. Our "evidence" from experimental economics, with reference to iterated trials, then suggests the subject's need to learn a "new" institution, but does not necessarily establish cognitive dissonance as an explanation for WTP-WTA differences in nonlaboratory experiments (Table 3.1) as we infer in III.4 and IV.C.

As to our observations of large WTP-WTA differences, this issue is addressed by Randall in Chapter VIII wherein he argues that, for a fairly wide range of contingent market designs, one can confidently expect that reported WTP and WTA measures will, respectively, understate and overstate an individual's true valuation. The generality of this conclusion (which we implicitly accepted in Chapters IV and VI) is challenged by Freeman as inconsistent with the "familiarity" issue discussed above in C-1-b: in instances where individuals lack accurate information regarding their preferences -- the CVM commodity takes the individual to preference orderings beyond the neighborhood of experienced consumption bundles -- individuals may make errors in any direction, i.e., WTP or WTA may be greater or less than values that would result from experience with the new commodity bundles. Along these lines, it is interesting to note that in

Bishop and Heberlein's Sandhill study (Chapter IX, Tables 9.2 and 9.3) hypothetical WTA values are less than cash offers ("true" valuations?) and WTP measures exceed cash offers; they also note large WTP-WTA differences in cash offers as well as offers involving hypothetical payment/compensation.

Kahneman strongly supports our "use WTP, not WTA" ROC, but first suggests that it be generalized and second, rationalizes the generalized ROC along different lines. His generalized ROC is: use the CVM only for commodities that have a "transactions structure"; do not use the CVM for commodities that have a "compensation structure." A "transactions structure" refers to a commodity-exchange context easily associated with voluntary exchange -- one pays for a commodity or action which makes him/her better off. A "compensation structure" refers to a commodity-exchange context wherein overtones of involuntary exchange are present -- how much you must be paid to accept more polluted air. The rationale for Kahneman's suggested ROC is his appeal to "prospect theory" which, in essence, assumes that individuals evaluate gains and losses differently; more specifically, it assumes that individuals value losses disproportionately higher than (identical) gains. Thus, one would expect a subject's valuation of a gain (WTP) to be substantively different from his/her valuation of a loss of identical magnitude (WTA).

We must confess that the link between Kahneman's rationale and his recommended ROC is not perfectly clear. One might appeal to prospect theory as a means for explaining why WTP and WTA measures should be expected to differ, but this would not argue for or against the preferability of one measure over another. It might argue, however, that one must use value functions based on WTP for valuing environmental improvements, but that a different value function, based on WTA measures, must be used in valuing (costing) environmental degradations; i.e., one cannot move toward the origin along a "benefit" curve. But this observation could apply with equal force to our conclusion that WTP, not WTA, measures be obtained via the CVM. Our rejection of WTA measures derived with the CVM is, upon close inspection, based on the argument that they are less "stable" than WTP measures; i.e., they are more affected by iterative trials, questionnaire design, etc. We do not make the case that cognitive dissonance, or other psychological/economic factors, are more or less relevant for WTP or WTA measures. Large differences observed between the two measures obtain in CVM studies, and that WTA measures are "high" may be inferred as a motivation for our recommended ROC.

Vernon Smith (Chapter XII.E) casts the WTP/WTA argument in a different light. He asks if we are not confusing WTA/WTP differences for the same individual with such differences among individuals. He notes that such differences among individuals, even if large, should not be disturbing since such differences provide the basis for exchanges -- large differences may simply imply a low volume in market trading. In terms of WTA-WTP differences for the same individual, Smith seemingly rejects the assumption of small income effects which underlies the Willig (1976) argument leading to approximate equality between WTP and WTA. His experiment demonstrates, first, that several subjects persistently reported WTA and WTP that were substantively different; secondly, his experiment demonstrates that despite differences in WTA and WTP values reported by individuals in the experiment, when such values are used in a market demand/supply context, the

resulting prediction of post-trade market-clearing prices is more accurate than predictions drawn from expected utility theory. Thus, Smith argues that empirical evidence belies the theoretical expectations of "equal" WTP and WTA for individuals -- note here the consonance of this observation with those of Kahneman -- but that in a market context such differences across individuals can result in accurate pre-trade predictions of actual (post-trade) prices (valuations) at which commodities are traded.

There are some particularly interesting implications of Vernon Smith's argument which warrant further examination. Consider the following data from Smith's experiment given in Figure 12.4.

Measure	Trial:		
	1	2	3
(a) Predicted price from the expected utility model	\$ 1.25	\$ 1.25	\$ 1.25
(b) Predicted price from WTA and WTP	1.25	1.43	1.48
(c) Actual, post-trading equilibrium price	1.30	1.51	1.52
(d) Sum of WTA	16.47	10.62	13.86
(e) Sum of WTP	12.42	10.80	12.24

Smith's experiment suggests a method for addressing accuracy/calibration questions related to CVM measures. For example, for a commodity which is exchanged in the market, a CVM study might be conducted which collects WTP and WTA measures from each subject. Demand (supply) curves are estimated from WTP (WTA) measures. Comparison of the resulting predicted price with actual market price has obvious implications for the accuracy of CVM estimates of value. Most importantly, Smith's experiment provides empirical weight for Kahneman's argument that benefits (the area under a WTP-demand curve) attributable to an environmental improvement may be expected to differ from costs (the area under a WTA-supply curve) for an environmental degradation. In this regard, the reader should note the different "areas" (sums) for WTP-benefits and WTA-costs implied from Smith's results given above, particularly values (d) and (e) for the first trial in Smith's experiment.

Related to Vernon Smith's argument is the point raised by Rosen (XIII.D). Rosen argues that WTP/WTA differences may in fact reflect "selectivity" i.e., populations from which WTP and WTA measures are taken are not homogeneous populations. In this regard, Rosen points to Brookshire et al.'s earthquake study: those living on 2 fault may well be expected to value earthquake risks differently from those who do not live on a fault.

Based on these interesting exchanges, it would appear to us that the following conclusions are relevant for the WTP/WTA issue. First, we agree

with Freeman and Bishop-Heberlein that a compelling case has yet to be made as to the general relationship between WTA and/or WTP measures and "true" valuations; certainly our attribution of such differences to cognitive dissonance is little more than an assertion. As is argued below, this implies the need for considerably more attention being given to the collection and analysis of psychological and attitudinal data in future CVM studies. Secondly, we agree with Freeman that the above-discussed "familiarity" issue is relevant for assessments of WTP/WTA differences; however, the little available empirical evidence does not support the notion that such differences are systematically related to the subject's familiarity with commodities. Referring to Table 3.2, WTA/WTP differences ranged from 2:1 to 5:1 in experiments involving private goods (goose permits in Bishop and Heberlein (1979) and a better-tasting substance in Coursey et al. (1983). Thirdly, we find Kahneman's "prospect theory" arguments to be, at a minimum, intuitively appealing, and certainly consistent with (if not supported by) considerable empirical findings. The notion that individuals value gains (from transactions structures) differently from losses (from compensation structures) may not, however, lead one to reject CVM applications to the estimation of WTA values; rather, it may suggest particular uses of WTP and WTA values: WTP for gains and WTA for losses. Finally, we concur with Bishop-Heberlein (Chapter IX) that the "burial" of WTA may be premature and that additional research is required which focuses on explanations of WTP-WTA differences. Meanwhile, it appears to us, our ROC "use WTP, not WTA" may serve as an operationally useful guideline for ongoing research with the CVM.

3. Attitudes vs. Intended Behavior. In Chapter V (Section E) the authors reviewed the "attitude versus intended behavior" issue raised by Bishop and Heberlein (1979 and 1983) which focused on the question: do CVM value measures reflect attitudes rather than intended behavior, and to what extent do attitudes correspond with intended behavior? Essentially, we adopted Randall et al.'s (1983) position that since CVM questions asked for intended behavior rather than attitudes, problems of correspondence between attitudes and behavior were likely minimized. We acknowledged, however, the relevance of Ajzen and Fishbein's (1977) design criteria for improving attitude-behavior correspondence (specific targets, actions, context and timing). As an aside, Bishop and Heberlein (Chapter IX) may have found our treatment of this subject to be uninformed or shallow, but in light of the major emphasis given results from psychological studies throughout Chapter V, we find ourselves nonplussed by their assertion of our "indifference and hostility" (Chapter IX, Section E.7) to the relevance of psychological research for economic inquiry. We confess, however, to understating the importance of attitude-behavior issues in psychology research.

Bishop and Heberlein's elaboration of the attitude-behavior issue in Chapter IX.E.7, is insightful, illuminating, and we believe, rich in its implications for the state of the arts of the CVM. Their major focus is on attitudes (as they relate to reported WTP) and behavior (actual payment of WTP) and the factors which result in close correspondence between the two. Attitudes are determined by the interaction of three components: cognition (dispassionate facts/beliefs), affectation (evaluative/emotional reactions to cognitive information) and intended behavior (intentional

"conclusions" derived from affective responses to cognitive information). Interaction between these three components is of primary importance; e.g., an affective change may motivate the individual to acquire more information (a cognitive change) which may then lead to a change in intended behavior. They argue, that a cash offer for a goose/deer license may elicit an affective response, and therefore a behavioral response, that is distinct from the affective response to a hypothetical offer -- witness their observed differences between valuations involving real and hypothetical payment. This analogy is consistent with Kahneman's arguments concerning WTP-WTA differences: WTA questions involving compensation structures elicit affective responses that differ from those elicited by WTP questions involving transactions structures.

Of primary interest are the factors which lead to close correspondence between attitudes and behavior. As an example in this regard, define AC (awareness of consequences) as a measureable manifestation of the cognitive component of attitudes vis-a-vis a CVM "commodity," and AR (acceptance of personal responsibility) as a measureable manifestation of the relevant affective component of attitudes. One can then define design and analytical criteria for assessing the probable correspondence between reported willingness to pay and what a subject might actually pay for a CVM commodity. Design criteria are those proposed by Ajzen and Fishbein (1977) to which we add questions related to AC and AR (see Bishop and Heberlein's examples in IX.E.7). In analytical terms, one's assessment of the probable correspondence between attitudes and behavior -- which relates to the probable accuracy of estimated values -- is based on the values of AR and AC variables. For the commodity in question, the greater is a subject's awareness of consequences (familiarity with the commodity?) and acceptance of personal responsibility, the greater is our expectation of close correspondence between attitudes and behavior (and, therefore, the more accurate the resulting measure of value).

As noted above, Bishop and Heberlein's elaboration of the attitudes-behavior issue allows for sharp focus on the need for attitudinal information for assessments of CVM results as well as for the types of information that would be useful in these regards. While not affecting the weight of their contribution, however, their discussions raise several questions of interest for our broad state of the arts assessment of the CVM. First, in operational terms, we simply note in passing the indexing task implied by their proposed criteria for correspondence between attitudes and behavior; e.g., what constitutes "high" values for AC or AR variables? Secondly, absent from their discussions is the relationship between attitude-behavior criteria and the other psychology-related issues discussed in Chapter V and reviewed by them. As an example, Bishop-Heberlein's discussion of the three interactive components of attitudes would seem to bear directly on the familiarity issue discussed above. If the cognitive component is empty -- subjects are unfamiliar with the commodity, or have little in the way of relevant facts/beliefs -- what might we expect in terms of affective responses and formulated behavioral intentions? A response to this question is implied in Kahneman's discussion of starting points (Chapter XII.C): subjects are simply incapable of assigning values to the commodity. Bishop-Heberlein's counterpart to this conclusion would seem to be: low AC values imply divergence between attitudes and behavior and thus (one supposes) inaccurate values.

A third question raised by Bishop and Heberlein's attitude-behavior discussions concerns the conflict between their position on the viability of estimating such things as option and existence values with the positions taken by us in Section VI.E and by Kahneman in XII.C. Appealing to familiarity/experience factors underlying our ROC's 1 and 2, we argue that one can expect a priori that such values must involve (using Freeman's model, Chapter IX) consumption bundles well beyond the neighborhood of bundles with which the subject has experience; thus, our rejection of uses of the CVM for estimating such values. Kahneman rejects the use of survey methods for valuing all but user values -- explicitly excluding option/preservation values -- in his discussion of "symbolic demand". Responses to questions related to ideological values, he argues, must reflect the subjects' hierarchy of values which tend to be injected into responses involving political or economic content. While acknowledging, first, that assessments of the validity of existence values via the CVM will not be easy and, secondly, that results from field experiments hold little promise for the use of the CVM in deriving such values, Bishop-Heberlein seemingly take the position that the CVM might indeed be used for estimating option or, particularly, existence values. The relative accuracy or meaningfulness of such measures would be assessed via analyses of the correlation between reported existence values and AC/AR variables. In their acid rain example, high existence values would imply (i) "high" awareness that acid rain damages will affect future generations (an AC variable) and (ii) a "high" indication that the subject feels personally responsible for reducing these effects (an AR variable; see IX.E.7).

In terms of the different positions concerning the use of the CVM for nonuser values described above, we should acknowledge possible exceptions to our conclusion that the familiarity/experience ROC's preclude the estimation of nonuser values; but we do not find Bishop-Heberlein's arguments (and the acid rain example) compelling in this regard. "High" AC values, which indicate familiarity with the acid rain problem, and "high" AR values simply do make their case: other values in the affectation "account" -- perceptions of how the subject is affected in a "user value" sense -- are relevant. At issue then is the subject's ability to differentiate between that part of his/her affective reaction to acid rain that is attributable to personal effects (a use value) and, generally, more altruistic affective reactions vis-a-vis future generations. Echoing Kahneman's notion of symbolic demand, it is this latter process, a process with which we expect the subject to have little experience, that we question. We would expect, a la Kahneman, that the sum of the user and nonuser parts will greatly exceed the subject's valuation of the whole.

4. Hypothetical Biases in the CVM: Conclusions In the authors' view, discussions at the Assessment Conference were particularly productive in giving perspective and context to the myriad issues concerning hypothetical bias discussed in Chapters III - VI. As noted in those Chapters, the potential for hypothetical bias in the CVM enters through the hypothetical nature of payment as well as the hypothetical commodity and the institution within which the commodity is exchanged -- the contingent market. We now ask, in light of the Assessment Conference, what is the state of the arts of the CVM in terms of the potential magnitude of-hypothetical biases?

In terms of hypothetical payment, we view the potential for related

biases with a great deal more equanimity than that suggested in the conclusions to Chapter V. In this regard, Mitchell and Carson's arguments as to the weakness of empirical results used by us in arriving at our more pessimistic conclusions are well made. The weight of the "incentives for accuracy" argument -must, at worst, be questioned in light of Vernon Smith's experiments, wherein WTP/WTB interrogations were "good" predictors of market outcomes, and the results from comparison studies wherein the CVM generated value estimates that were remarkably close to estimates derived from indirect market methods (holding the question of the accuracy of any method aside, for the moment). We concur with Arrow's observation that hypothetical/real payment differences may not be as serious as one might fear: well designed survey instruments wherein the exchange setting is "pseudo-real" may indeed elicit real-like results. This is not to argue that incentives/hypothetical payment issues are not relevant; it is to argue that, first, the jury is still out -- it remains an open issue -- and, second, that some promise exists for structuring CVM instruments in ways that mitigate, if not eliminate, the magnitude of payment bias.

Within the rubric of "hypothetical bias," we find the most prominent source of bias to arise in instances wherein the CVM commodity, within a contingent exchange setting, is largely unfamiliar to the subject -- the subject has no experience in viewing the commodity within the context of trade-offs. In Freeman's terms, the effect of the CVM is to move the individual to areas of his preference orderings that are far removed from neighborhoods of consumption bundles with which the subject is familiar. Our lack of models concerning subjects' behavior in the CVM setting notwithstanding, we see in Freeman's rudimentary modeling efforts, as well as in Kahneman's notion of symbolic demand and Bishop-Heberlein's discussions of the roles of attitudes, the bases for reasserting our contention that, for state of the arts applications of the CVM, (i) participants in the CVM must understand (be familiar with) the commodity to be valued (our ROC Number 1) and (ii) subjects must have had (or be allowed to obtain) prior valuation and choice experience with respect to consumption levels of the commodity (our ROC Number 2).

In terms of learning issues, final state of the arts conclusions concerning the efficacy of iterative bidding processes and laboratory methods/techniques for applications of the CVM must be softened considerably from the tone of earlier conclusions offered in Chapters III - VI. We find impressive the substantive effect on bids that result from the iterative bidding process in studies involving, not just the small samples of concern to Mitchell and Carson, but large sample sizes. In our view, iterative bidding does result in substantively higher bids. Iterative effects notwithstanding, Mitchell and Carson, as well as Bishop and Heberlein, are obviously correct in pointing to the lack of evidence that would support (or reject) the attribution of such effects to the preference research processes as asserted by us in Chapters III - V; moreover, we must acknowledge the substance of Bishop and Heberlein's observation that the parallel between the iterative bidding process and the iterative valuation trials used in laboratory experiments, implied by our discussions in IV, is without obvious substance. Nor, it seems fair to say, has the attribution of iterative bidding effects to Mitchell and Carson's "bullying" or "social awkwardness" motives been established. Thus, all that can be said at this point in time is that iterative bidding rather consistently results in higher CVM

valuations, but we are unable to explain such differences.

Bishop and Heberlein's lament that economists involved in CVM research are woefully ignorant of research results in the related, and certainly relevant, field of psychology extends with equal force to economists' general ignorance (until only very recently) of developments in experimental economics; the authors concede their general ignorance in this area prior to the development of this book. As the novice enters the literature of experimental economics, he/she must be struck with the impressive developments made in that field which relate directly to the most perplexing questions facing the CVM practitioner: how does one establish incentive structures; how do subjects learn; how does one elicit preference revelation? The real "lessons" from experimental economics of unquestionable importance for the development of the CVM are found in two principal areas. First, laboratory methods can provide us with a relatively inexpensive and efficient method for conducting experiments concerning design and conceptual questions of relevance for the CVM; examples in these regards are questions concerning strategic bias, WTP-WTA differences, effects of psychological variables on subject valuations, etc. Secondly, and of particular importance, developments in experimental economics may be provocative -- challenging -- to CVM researchers in terms of stimulating new and imaginative lines of inquiry concerning persistent problems encountered with the method. In these regards, the issue is not, for example, whether or not the Vickery Second Price Auction per se will "work" in applications of the CVM; rather, the issue is: can the CVM be structured so as to better provide incentives for true revelations of preferences (as an interesting initial effort in this regard, see Bishop and Heberlein's experiments with a Fifth Price Auction in Chapter IX). As another example, can we (should we) be experimenting with repeated visits (repeated "trials") with CVM subjects, with questions designed to help them learn incentive-compatible behavior vis-a-vis a contingent market?

Thus, lessons from experimental economics are clearly relevant for our State Of the arts assessment of the CVM: they indicate the lack of substantial progress made in the method's development in important areas Concerning subjects' learning/understanding of incentive structures. Such lessons are not, however, a panacea for resolving the problems of the CVM. Earlier-noted comments by Conference participants concerning our over-emphasis on the ready transferability of methods/techniques used in experimental economics to applications of the CVM for valuing public goods are well made, as are the reminders by Arrow and by Bishop-Heberlein of the important role of field experiments for improving the state of the arts Of the CVM.

Turning now to the WTP-WTA issue, relevant state of the arts conclusions were suggested in the closing paragraphs of Section C.2 above. V. Kerry Smith's call for theoretical inquiry as to subjects' behavior in the contingent market setting is particularly appropriate for efforts to explain WTP-WTA differences. In this regard, see the contrast between Randall's theoretical model, which relies on subjects' perception that their responses influence policy, wherein WTP (WTA) understates (overstates) "true" valuations, and Bishop-Heberlein's contrary evidence as well as Freeman's model which suggests that, for "unfamiliar" commodities, WTP or

WTA relationships to true valuations cannot be determined a priori. While we find compelling, on deductive as well as intuitive grounds, Kahneman's argument that subjects value losses differently than gains, we are concerned with the fact that WTA measures appear to vary much more than WTP measures in response to such things as iterative trials. Thus, in operational terms, i.e., as we await results from further theoretical and empirical research concerning this question, we maintain our conclusion suggested in Chapter VI which states that WTP, not WTA, measures should be estimated with the CVM.

Finally, the state of the arts of the CVM in terms of our appreciation of the attitude-behavior issue is, in our view, greatly enhanced by Bishop-Heberlein's discussions in Chapter IX. Means by which the accuracy of CVM measures, in terms of the correspondence between attitudes and actual behavior underlying reported willingness to pay, are directly implied by the interactive relationships between attitudinal components and behavior. While implementation problems remain for resolution, one can see in Bishop-Heberlein's exposition the essential framework for deriving empirical measures for cognitive and affective components of attitudes and, at least conceptually, their use in deriving indices of attitude-behavior correspondence.

D. THE ACCURACY OF CVM MEASURES OF VALUE

1. Overview of the "accuracy" issue. Recurring throughout Part I of this book, as well as throughout Conference papers and discussions, is reference to a subject's "true" valuation of a public good such as an environmental change. Thus, our standard for accuracy in values derived from the CVM is a subject's reported valuation that reflects a "true" revelation of preferences vis-a-vis the CVM commodity. In this regard, our appeal to market institutions as a framework whose structure we hope to simulate in the process of applying the CVM is motivated by our desire to capture, in applications of the CVM, the incentives for preference revelation that our theories lead us to expect from a market context. In the market context, individuals must introspectively balance the utilities foregone as a result of paying for a good with the utilities gained from acquiring the good; to this end, he/she must, however "completely," search his/her preferences for the good in question vis-a-vis all other possible goods and their prices (relative to his/her income). Thus as has been extensively argued above, the importance for assessments of the CVM of such themes as the subject's familiarity with a commodity (for the preference "search", or research process) and the credibility of payment and payment modes to the subject (for meaningful subjective assessments of implied trade-offs).

In these regards, we must reiterate our earlier-noted concern with V. Kerry Smith's interpretation of our ROC's related to these themes as requiring that the value derived in the CVM be the subjects' estimation of market outcomes as opposed to the subjects' preference revelations; ROC's per se are discussed below. ROC-1 requires that the subject have some familiarity with the CVM commodity and ROC-2 requires some choice experience, direct or indirect, with respect to consumption levels of the CVM commodity. These conditions then loosely require that, as in Freeman's arguments, the consumption bundles (including the CVM commodity) that the subject is hypothetically evaluating are within neighborhoods of consumption bundles with which he/she has had experience. Thus, our concern with accurate revelations of preferences leads us to require that choice setting which is analogous to a market setting, and which is consistent with the expectation that the subject is capable of meaningful searches of preferences. To require an "informed" choice setting does not, in our view, imply that the CVM application must then elicit the subjects' introspective estimate of solutions of a hedonic market.

Given that our standard for CVM values is the true revelation of a subject's preferences, the primary question becomes: how do we measure that standard? Obviously, if we had a "true" value, assessments of the accuracy of CVM measures vis-a-vis this standard would be straight-forward. The state of the arts relevant to such measures is such that, aside from limited results from laboratory and field experiments involving private goods, these measures are not available. Therefore, in Chapter VI the question of the accuracy of CVM values is addressed in the following indirect and inferential manner. First, we note the literature that suggests that, for ordinary demand studies based on "hard" market data, estimates may involve errors (the range for accuracy might be) on the order of $\pm 50\%$ or more. V. Kerry Smith (Chapter XI) expands on this argument, arguing that much of economist's "hard" data may be subject to the same type

of criticism concerning, e.g., hypothetical and reporting biases as those leveled at the CVM. Such errors are generally attributable to such things as assumptions concerning the distribution of error terms and functional forms. Secondly, from these data we then infer that econometric value estimates based on indirect market methods would involve ranges of error no less than those in ordinary demand analyses, i.e., one can argue, at most, that indirect market methods yield value estimates which would encompass "true" values within the range $\pm 50\%$. Thirdly, appealing to the concept of "reference accuracy," we note that received studies demonstrate that biases associated with starting points, payment vehicles, information and iterative bidding could result in errors as large as $\pm 50\%$ in CVM studies.

In retrospect, we might well have stopped our arguments here: available evidence suggests that either the CVM or indirect market methods may yield estimates of "true" preference revealing values within a range no better than $\pm 50\%$. We carried these arguments a step further, however, in addressing the following question. Noting -- uncritically, it must be acknowledged -- cited instances wherein economists quite comfortably impute accuracy to market-based estimates of value, we implicitly construct the following strawman: suppose that indirect market methods yield accurate results -- "accurate" within the range $\pm 50\%$; are value estimates from indirect market and contingent valuation methods different? We continue by positing that if they are not different, then the accuracy of indirect market values implies the accuracy of CVM values. Referring to the fifteen CVM-Indirect Market study comparisons given in Table 6.12, and noting that ranges ($\pm 50\%$) for accuracy of CVM values overlap with those for indirect market methods in 13 comparisons, we then conclude that, for commodities which are amenable to application of indirect market methods (a caveat then used to form ROC's), the CVM may yield value estimates that are as accurate as (the assumed accurate) values derived from indirect market methods. It should be noted that any specification for the magnitude of errors associated with the use of the CVM is premature at this time. We choose $\pm 50\%$, as a means for focusing attention on what is, in our view, an interesting approach for assessing the accuracy of CVM measures.

In many ways our discussions of accuracy achieved their intended purposes: they certainly received the attention of Conference participants; most importantly, they succeeded in initiating a dialogue focused on how future research might address calibration and accuracy issues. Constructive criticism of our discussions of accuracy offered by Conference participants may be seen as involving the following three sets of issues.

2. What is Accuracy? The first set of issues involves the question as posed by Arrow: what do we mean by "accuracy" and what level of accuracy is it reasonable to expect from applications of the CVM? In response to these questions, Arrow offers four observations: (i) referring to hypothetical issues, the reality with which economics (and other social sciences) deal, involves counter-factual lines of deduction -- statements comparing actions with states that "would" hold, but in fact do not. Our concern is with questions of the form: what would we do if reality were marginally different (e.g., if income were one unit higher)? In virtually, all cases, the "truth" relevant for these questions can never really be known; (ii) inaccuracies in real-world efforts to estimate individual

preferences via demand analyses based on "hard" data are probably best seen in the fact that half of the "new" products put on the market fail. (iii) our colleagues in medical and engineering sciences consider, as a matter of course, estimates producing errors on the order of one to ten (one order of magnitude, see VI.D) to be normal; (iv) therefore, it is not clear that we should be disturbed if our value estimates are thought to be within $\pm 50\%$ of true values, or $\pm 100\%$. Ranges of error of 3:1 or 5:1 may pale in significance when compared to those reflecting technical ignorance in most environmental fields.

V. Kerry Smith also stresses Arrow's point that we can never know "true" valuations. Indeed, in our general scientific inquiry we never prove hypotheses, we fail to reject them. Arrow's reminder of the limitations of "hard" data vis-a-vis their use in estimating value is expanded by Smith along interesting and provocative lines. In Table 11.1 (Chapter XI), he demonstrates the potential for strategic and hypothetical biases (broadly defined) in various sources which are generally thought to produce "hard" -- accurate -- data.

As an aside, we are compelled to note the contrast between Arrow's and Smith's arguments and the framework for considering the question of accuracy offered by Freeman (Chapter X.E). Define B as a subject's response to a CVM question and assume that B is a random variable with mean, B' ; B^* is the individual's true valuation. Freeman's suggested approach for analyses of accuracy is then one which focuses on $B' - B^*$ and on the variance of $e = B - B'$. He distinguishes between "biases" -- $B' - B^*$ differences attributable to starting points, information, etc. (the topic of Section A above) -- and random errors reflected in $B - B'$ differences, where random errors result from the hypothetical character of the CVM (the substance of Section C above). With biases eliminated by questionnaire design, and assuming that e is normally distributed with zero mean, large samples (which would result in $e = 0$) may result in $B' = B^*$. In the light of our earlier discussions, the application of Freeman's approach involves two major questions, satisfactory responses to which elude the authors. First, on what basis does one argue in a compelling way that the many sources for hypothetical biases are random and, particularly, normally distributed with zero mean? Most importantly, and directly related to Arrow's and Smith's arguments, how does one divine the "truth" -- whence comes B^* which critically serves as the basis for assessing the effectiveness of questionnaire design in eliminating "biases"? In the scientific literature, the concept of measurement accuracy rejects the notion that "true" valuations can be known, the result of which is a focus on removing demonstrable errors.

Finally, Bishop-Heberlein's arguments have implications for the question: what is accuracy? In terms of the accuracy of values derived from the CVM, their discussions would seem to imply that accuracy turns on the correspondence between attitudes and behavior, wherein such correspondence might be in some sense measured by Azjen-Fishbein criteria (vis-a-vis questionnaire design) and by cognitive and affective variables. In passing, we note their second (tongue-in-cheek) criterion for accuracy which was suggested at the Assessment Conference: "good enough for government work", which might (quasi-seriously) be taken to mean that order of magnitude estimates may be regarded as "accurate" for some

applications of the CVM.

3. Reference Accuracy and public good values. In our efforts to couch the accuracy issue in terms of "Reference Accuracy" -- accuracy is defined in terms of biases resulting from deviations from Reference Operating Conditions -- the approach per se was well received by Conference participants; our exposition of a numerical application of the approach was not. In this latter regard, our $\pm 50\%$ argument was seen as "weak" by Freeman, as being "ad hoc" by Rosen, and unconvincing by Mitchell and Carson. Referring to the CVM as well as indirect market methods, V. Kerry Smith questions the extent to which any error range can be imputed to estimated value measures given the present state of our knowledge. The basis for much of the expected criticism of our (no better than) $\pm 50\%$ reference accuracy range for CVM measures reflects several related arguments which, we of course concede, are well made. Mitchell-Carson, Bishop-Heberlein and Rosen point to the fact that well-designed CVM studies need not include biases resulting from starting points, payment vehicles, information and/or iterative bidding. Indeed, our discussions above in Sections B and C suggest that payment cards can be structured so as to mitigate or eliminate starting point biases; payment vehicle bias may be a misnomer -- mode of payment may be inextricable from the commodity; and, particularly for "familiar" goods, information issues may be amenable to control by questionnaire design. Thus, these individuals argue, demand studies using the CVM (or indirect market methods) are not of equal quality, as is implied by our general statement that reference accuracy for the CVM may be no better than $\pm 50\%$. To these arguments Mitchell and Carson add the observation that sampling errors, discussions of which were excluded from our assessments of the CVM, must also be considered -- sampling errors alone could result in errors of $\pm 50\%$.

4. The need for accuracy or calibration research. In the physical sciences, Reference Accuracy, based on ROC's, is the accepted practice for evaluating the precision of instruments for measurement. Generally speaking, Conference participants were supportive of our efforts in Chapter VI which were designed to initiate thought and research concerning means by which ROC's might be defined and by which we might measure the error implications of CVM applications wherein one or more of the ROC's are not satisfied. Thus, Arrow calls for more field and laboratory experiments designed to establish conditions under which reasonably defined accuracy in the CVM might obtain, a call echoed by Rosen who, in addition, feels that replications of CVM studies might be useful in these regards. Bishop-Heberlein appeal for research designed to calibrate errors with the extent to which ROC's are satisfied. V. Kerry Smith's insistence on the need for modeling efforts is joined with his observation of our lack of knowledge as to how violations of ROC's affect subjects' valuations.

Of course, the need for standards against which the accuracy of CVM values might be assessed underlies our suggested ROC's. Given the critical importance of ROC's for the use of Reference Accuracy, attention is now turned to an evaluation of those conditions.

5. The Reference Operating Conditions. There are at least two requirements for estimation and use of Reference Accuracy for the CVM: the specification of Reference Operating Conditions -- the conditions or circumstances which limit the accuracy of a measurement tool; and the

magnitude of errors which result from failure to satisfy any given ROC.

Consider, first, the problem of specifying ROC's relevant for the CVM. That ours is not the last word on ROC's relevant for the CVM is made clear by ROC's explicitly or implicitly suggested by Conference participants. Referring to Table 13.1, ROC's 1 through 4 are those suggested by us in Table 6.13; ROC Number 8 was implied in our discussions of the $\pm 50\%$ Reference Accuracy range for the CVM but, for reasons which now escape us, was not explicitly included as an ROC. ROC's 5-7 are those suggested by Kahneman -- note the overlap with ROC's 4 and 5. Mitchell and Carson suggest, based on referenda and psychological research, ROC 9 (and concur with ROC's 1, 3 and 4). A choice for an ROC Number 10 is implied by the apparently contradictory positions of Randall, who would require subjects to view the CVM process as a real opportunity to influence policy, and Arrow, Freeman (1979) and, we should add, Rosen, who would view a subject's perception of the CVM process in such a real, nonhypothetical way as possibly inviting strategic responses. Finally, Bishop-Heberlein's discussions imply ROC 11.

TABLE 13.1

ALTERNATIVE REFERENCE OPERATING CONDITIONS

Reference Operating Condition	Measurement Error When ROC is not Satisfied
1. Subjects must understand, be familiar with, the commodity to be valued.	?
2. Subjects must have had (or be allowed to obtain) prior valuation and choice experience with respect to consumption levels of the commodity.	?
3. There must be little uncertainty.	?
4. WTP, not WTA, measures are elicited.	? \pm 300%
5. (Kahneman) Valuations must involve transaction structures, not compensation structures.	? \pm 300%
6. (Kahneman) CVM values obtained must relate to use, with minimum ideological content.	?
7. (Kahneman) Payment vehicles must be well defined and credible vis-a-vis the CVM the CVM commodity; values obtained with one vehicle may not be interpretatively "transferred" to those which we would obtain with other vehicles.	?
8. CVM applications must involve:	
(i) No basis for starting points or anchoring;	?
(ii) "appropriate" information concerning the commodity and the valuation process;	?
(iii) initial, noniterated valuations.	?
9. (Mitchell-Carson, from referenda/psychological research):	
(i) Subjects must be given as simple a choice as possible;	?
(ii) outliers should not unduly influence research;	?
(iii) subjects should be permitted to abstain from the valuation process.	?

10. (Implied by Randall, Chapter VIII): Subjects ?
must view the CVM process as a meaningful
opportunity to influence policy via their
responses;
or
(Arrow, Rosen and Freeman, 1979): Subjects ?
must view questions as being sufficiently
hypothetical so as not to provide incentives
for strategic behavior.
11. (Bishop-Heberlein):
(i) Azjen-Fishbein criteria for the structure ?
of valuation questions must be satisfied.
(ii) "close" correspondence between attitudes ?
and behavior is required.

It must be acknowledged that the rationale for including any of the ROC's in Table 13.1, as well as the rationale for excluding other possible ROC's, is weak or nonexistent at this point in time. For example, our suggested ROC's 1 and 2 are justified by, first, the "familiarity" argument and secondly, our observation that in several comparison studies, $\pm 50\%$ accuracy ranges for CVM values overlap with $\pm 50\%$ ranges for indirect market methods for valuations of commodities which we assert are commodities with which subjects are probably familiar and have some degree of indirect market experience. Obviously, neither argument is immune to challenge. As a further example, in Section C.4 above we argue for the abandonment of the "information bias" rubric (ROC 8 (ii)). As a final example, we note that at this stage of the state of the arts, we are unable to even give precise definitions for many of the limits on CVM measures that we believe to be important; e.g., in 9(i), what is a "simple" choice?; in ROC 10, what is a "meaningful opportunity" or a "sufficiently hypothetical" choice?

Thus it is hoped that the combined discussions in this book concerning the potential role of ROC's in providing means by which ranges of Reference Accuracy may be attributed to CVM measures will provoke imaginative thinking and research relevant to the specification of precise and defensible ROC's; in any state of the arts assessment, of course, the immediately preceding discussions establish the infant stage of this process at this point in time.

As is obvious from Table 13.1, while we at least can see a place to begin in terms of specifying ROC's, our knowledge is virtually nil in terms of the error implications of not satisfying an ROC. Referring to ROC 8 in Table 13.1, Rowe and Chestnut's (1980) error estimates can be of very limited usefulness for our purposes given our inability to assess the quality of studies used in their samples vis-a-vis other relevant ROC's. Of course, this virtual void in our knowledge is the motivation for the insistence on "calibration" research by almost all of the participants (see, particularly, the Comments by Arrow and Rosen in Chapter XII, and those by Bishp-Heberlein (Chapter IX) and by V. Kerry Smith (Chapter XI)).

E. THE STATE OF THE ARTS OF THE CONTINGENT VALUATION METHOD

In Chapter I we noted the need for a "reflective pause" in CVM research wherein concerned researchers can take stock of the progress that has been made in the development of the method, and of the major issues which require resolution for further developments. The need for such a pause was made manifest by our review of the myriad "criticisms" of the CVM, all of which pointed to the disarray and confusion amongst CVM researchers attributable to two central facts. First, there has been a lack of consensus among researchers as to the priority issues and hypotheses that warrant empirical focus. Research efforts appeared scattered and diffuse as we repeatedly addressed asserted "biases" in the CVM (e.g., starting point, information, vehicle biases, etc.) in the "heuristic" manner described in Chapter III, with seemingly but one basis for accepting or rejecting a "bias": some ill-defined "preponderance of evidence." In large part, this lack of a well-defined, prioritized research agenda for the CVM reflects the ad hoc, "chemistry set" approach to CVM research noted by V. Kerry Smith, Bishop-Heberlein, and other Conference participants. Empirical applications of the CVM have outstripped intellectual inquiry -- via formal models or otherwise -- as to how individuals may behave within contingent market settings and implications for questionnaire design and implementation practices. Secondly, following perhaps from the preceding observations, CVM researchers have been apologetic, or defensive, vis-a-vis the "rest of the profession" due to the pervasive feeling that interrogated responses by individuals to hypothetical propositions must be, at best, inferior to "hard" market data or, at worst, off-the-cuff attitudinal indications which might also be expected to reflect efforts by individuals to manipulate the survey to their selfish ends.

The difficulties involved in efforts to provide some state of the arts context for the controversies surrounding the viability of the CVM for estimating values for public goods are made manifest by the assessment process seen in Parts I and II of this book. Thus, many of the positions and conclusions presented by us in Part I were altered or retracted in this Chapter as a result of the focused dialogue concerning priority issues in CVM research between the authors, four other prominent CVM researchers, and leading economics and psychology scholars whose interest in public goods valuation is a step removed from the CVM per se. Of course, the reader will judge the success of this process in providing a state of the arts context for the CVM. In this regard, our general view of this context is described as follows:

1. The CVM Without Apology. It is surely time for replacing apologies for the CVM with a positive research agenda to be described below. AS a first step in this direction, we must eschew the joys of self-flagellation over our lack of knowledge of the "truth": we don't and won't know it, nor will our colleagues in the "rest of the profession" vis-a-vis their value estimates, nor will scientists in other disciplines. Following Arrow's exhortations, we must directly address the question, what is accuracy, and then look to calibration methods which provide us with a means to achieve accuracy levels that are reasonable and consistent with those levels obtained in other areas of economics and in other disciplines.

What is accuracy in a CVM estimate? It is a subject's valuation of a commodity which "reasonably" reveals his/her preference for the commodity. What does "reasonable" mean? "Reasonableness" is established by criteria -- Reference Operating Conditions -- which allow us to measure the magnitude of probable errors in any given application of the CVM. Thus, whether resulting ranges for Reference Accuracy associated with applications of the CVM are never better than $\pm 50\%$ or $\pm 500\%$, our focus is on defining the reference accuracy range. As with any other estimates, the "usefulness" of estimates with any range of error is determined by the purposes to which the estimates are to be put.

2. Conclusions concerning accuracy. While perhaps useful in pointing to needed research, the above is little more than a definition of accuracy. Given, as was argued above in Section D, that efforts to develop ROC's relevant to the CVM have just begun, and that we are almost totally ignorant of the error implications associated with the few ROC's that seem palatable at this time, must we then agree with V. Kerry Smith's judgement (Chapter XI) that no conclusions about the accuracy of CVM measures can be drawn based on research accomplished to date? We think not. At this point of reflective pause in the development of the CVM, one fails to see implications for the accuracy of CVM measures from received research only if one's view of "acceptable" implications is limited to evidence that demonstrates some degree of precision -- narrow ranges of error. This is to say that while we cannot build the case for ranges of Reference Accuracy for the CVM of magnitudes that would make CVM value estimates of practical use in many cases, at this point in the method's development a "useful conclusion" in the sense of V. Kerry Smith's assertion might well be that the method produces order of magnitude estimates -- but we think one can argue that error ranges are much smaller.

Before continuing this argument, it is relevant for our purposes to recall V. Kerry Smith's demonstration (Chapter XI) of the wide range of potential for hypothetical and reporting errors in "hard" data commonly used, without apology, in economic analyses. Such data are seemingly accepted in total ignorance of ROC's relevant for their collection and the resulting ranges of Reference Accuracy. This observation, when combined with Coursey and Nyquist's findings of potential errors in ordinary demand analysis and Mitchell-Carson's general comments regarding sampling errors, should serve -- to paraphrase Freeman (Chapter X.E) -- as a chilling reminder of the limitations of empirical analysis/models in most areas of economic analysis. It seems fair to say that, in the general economics literature, questions of accuracy are not prominent. This is not to suggest a nihilistic approach to CVM research: the whole world is wrong (inaccurate), so why should we be concerned with accuracy. We mean to suggest the perspective: economists' typical preoccupation with such things as standard errors, etc., may have misled us into viewing value estimates as "precise" in terms of narrow error ranges, $\pm 5\%$, 10% or even 20% . Couched in the broader terms of Reference Accuracy, such "precision" in general economic value estimates may quickly dissipate. Again, that such broader views of accuracy are generally ignored in economics is made manifest by V. Kerry Smith's provocative discussion in XI.B.

Returning to our discussion of what one can conclude regarding the

accuracy of CVM measures, we begin by recalling an earlier discussion of the "truth". We do not and will not know it. But something analogous to "truth" may be attributed to values derived from, as examples, actual cash trades in Bishop-Heberlein's Sandhill study and in Vernon Smith's laboratory experiments. Eschewing arguments as to how Bishop-Heberlein's auction formats might have been improved in one way or another, their cash offers/payments are certainly the "truth" vis-a-vis preference revelation in the sense that folks clearly paid (were paid) for a well-defined commodity and then used the commodity. For the limited, most likely nonequilibrium, "simulated" market used by them, we can surely attribute preference revelations to these values. The differences between mean cash and CVM WTA values was roughly 42%; between cash and CVM WTP values, differences ranged from about 38% to 124% across their four auction formats (Tables 9.2 and 9.3). Do these differences imply nothing vis-a-vis conclusions as to the accuracy of CVM measures? If accuracy is viewed as involving "small" ranges for Reference Accuracy, one would lament the "large" differences, as do Bishop and Heberlein, and concur with V. Kerry Smith that nothing (positive) can be concluded. If orders of magnitude are relevant, one might find Bishop-Heberlein's results startling: CVM and cash offers are virtually the same (see Figure 6.1). Our colleagues in environmental engineering may well envy such accuracy. In these regards, we note Bishop-Heberlein's later "surprise" (IX.F) at how well the CVM does work -- cash-CVM differences were not "outrageous".

Questions of the transferability of laboratory results to real-world conditions aside, hypothetical responses in Vernon Smith's experiments were consistently within 10% of actual market outcomes. In the Coursey et al. laboratory experiment (Figure 4.1), differences between values derived from final Vickery auctions and hypothetical questions were less than 20% for WTP and approximately 100% for WTA. The central point in all of this is apparent, however. In terms of the standard for comparisons of CVM values, we can continually argue as to how well preference revelations are made manifest by Bishop-Heberlein's cash offers, Vernon Smith's securities values, Coursey et al.'s measures related to tasting sucrose octa-acetate, or, moving to public goods, TCM and HPM values derived by the eight sets of authors given in Table 6.12. But however well any of these measures reflect meaningful revelations of preferences by individuals, every piece of evidence that we have demonstrates that the CVM yields value estimates that are indistinguishable from those standards in order of magnitude terms. Indeed, and herein lies the relevance of our $\pm 50\%$ arguments, in most instances CVM values are within $\pm 50\%$ of values derived from alternative methods for estimating preference revealed values.

3. Final Remarks. Thus, our final (c.f. our stronger, pre-Conference, reservations in Chapters I - VI, ad passim) assessment of the state of the arts of the CVM is generally positive. We find impressive the accuracy of CVM measures inferred by the available evidence at this stage of the method's development. We find encouragement in the Conference results, Particularly those reported by Arrow, Kahneman, and Bishop-Heberlein, which suggest that breaking the "hypothetical barrier" in the CVM may not be as hopeless as we and others earlier believed.

"Promise" is not "performance," however, and our assessments given above refer only to the potential promise of the CVM as a viable method for

estimating values for public goods. The realization of that promise implies real challenges for theoretical and empirical research for those involved with the method's further development. In concluding this book, we now focus attention on critical issues for any research agenda which are relevant for guiding future CVM research.

F. CRITICAL ISSUES FOR FUTURE CVM RESEARCH

In the most general terms, it must be hoped that greater focus can be achieved in future research with the CVM. Both Bishop-Heberlein and V. Kerry Smith emphasize the ad hoc character of the bulk of CVM research to date -- a characterization aptly described by Bishop-Heberlein as reflecting a "chemistry set" approach. To a large extent, the ad hoc quality of CVM research has resulted from the emphasis or priority given empirical results -- necessitated in many cases by data needs of the entities providing research funding -- as opposed to theoretical and design issues. Results from this empirical emphasis are made manifest by the profession's preoccupation, without resolution, with such operational "biases" as starting point, information and vehicle issues as noted in Chapter III, issues some of which, upon reflection by Conference participants, may now be viewed as not implying biases per se but rather implying limits on questionnaire design and the manner in which CVM values are interpreted. Thus, the first critical issue for future CVM research is the metaphorical realignment of the empirical cart and the theoretical horse. There is a critical need for modeling efforts focused on individual behavior in contingent market settings which may serve as a basis for formulating hypotheses for empirical testing. This need for modeling efforts underlies virtually all of the additional issues for further CVM research discussed below.

A second critical issue for future research involves the specification and measurement of Reference Accuracy for CVM measures. In this regard, imaginative and innovative thought is required for defining relevant ROC's (e.g. Table 13.1 above) and for calibrating errors with deviations from ROC's. Thus, we must ask questions exemplified by: What is "familiarity" or "experience" vis-a-vis a CVM commodity; what is "uncertainty" and what constitutes "ideological content"; what variables may perform best as measures of cognition and/or affectation and how are attitudinal variables calibrated with measures of attitude-behavior correspondence; how can we better structure value questions so as to enhance a priori our expectations that preference revelations are obtained which are at least consonant with incentive-compatible revelations in market contexts? In addressing these issues we will need to profit from and exploit the lessons learned in laboratory and field experiments, as well as in research in other disciplines.

A final critical issue for future CVM research involves our need to resolve the "incentives" question. In this regard, our concern extends beyond the hypothetical payment question. We concur with Arrow's suggestion that question settings that are sufficiently pseudo-real may be expected to result in satisfactorily pseudo-real responses and we are not convinced as to the extent to which one can distinguish between payment effects and those attributed to familiarity and experience questions. Of interest in these regards is the threads of an argument, seen implicitly in Randall's paper, as well as in Kahneman's Comments, and explicitly in Mitchell-Carson's paper (Appendix), that valuations of contingent changes in provision levels of public goods might be better obtained via processes which attempt to simulate results from institutions other than the market institution. Their examples specifically suggest the referendum institution. In terms of familiarity and experience, the provision of public goods via

reliance on market-like transactions valuations is, at best, tenuous vis-a-vis the referendum process which is actually used in this regard. Some sort of preference revelation must surely be inferred by the act of an individual's signing a petition which requests a public/social action which the individual generally knows will result in his/her payment of higher taxes. Thus, a la Randall, the subject may indeed be motivated by the opportunity to influence policy. Whether such motivation would lead to "strategic" signings of a cost-specific referendum is an important empirical question. Here we simply note the potential appeal for such a variation in CVM applications in dealing with many of the sources of familiarity/experience problems, when market analogies are used in the CVM and its possible use in resolving (or re-casting) the incentives problem.

APPENDIX - Chapter XIII
SOME COMMENTS ON THE STATE OF THE ARTS ASSESSMENT OF THE
CONTINGENT VALUATION METHOD DRAFT REPORT*

Robert Cameron Mitchell

Resources for the Future

and

Richard T. Carson

Resources for the Future

and

The University of California (Berkeley)

A. ASPECTS OF THE USE OF THE BIDDING GAME FORMAT (Chapter III)

1. Starting Point Bias. In our view, the evidence for starting point bias is far stronger than the draft report's review of this literature (p. 59ff) suggests. Although the authors appear to recommend against the use of bidding games at the point of eliciting the initial bid, presumably because they feel that starting point bias is a real problem, other readers might review the evidence presented in the report and conclude that starting point bias is not a problem with the bidding game format. In what follows, we present a critique of the report's literature review on this topic.

In the first place, the literature review offered in the report includes various items which are extraneous to the issue of starting point (SP) bias but which, nevertheless, appear to be offered as evidence against SP bias.^{1/} Secondly, there are several other studies not cited in the draft report whose findings support the notion of SP bias. These include our reinterpretation of Greenley, Walsh and Young's water benefits study (Mitchell and Carson, forthcoming); the study by Thompson and Roberts (1983) of recreation values for offshore oil platforms which shows a strong effect in a well-designed test; and a forthcoming paper by Boyle, Bishop and Welsh which also shows a strong effect in a well-designed test.

Third, some of the previous tests for SP bias, which are interpreted in the report as showing no SP bias, are potentially flawed because of the low power of the tests. It is well recognized that sample size decisions should take into account the size needed to detect a specific difference with a specified power.^{2/} Hypothesis testing on small samples which have fairly high coefficients of variation face the problem of accepting a finding of "no difference" a large percent of the time when in fact a difference of as much as twenty five percent may be the case.^{3/} Given the very small samples used in the Los Angeles tests for starting point bias (p. 61), the likelihood of finding a difference at the .05 or even .10 level was very small. That a few of the tests in that study did find differences should be viewed as a potential sign of strong starting point bias rather than as evidence that it is rarely found. For the same reason, Desvousges, Smith and McGivney (pp. 64-5) were unable to positively assert that starting points of \$25 and \$125 caused bias in their study despite the fact that the difference between their means is large and in the predicted

direction.4/

2. Iterative format. At various points the report emphasizes the importance of using an iterative format in CV studies. The grounds offered for this recommendation are several. (i) The finding (p. 67) that an initial payment card bid is raised significantly (despite small sample sizes) when respondents are told the amount they originally offered may not be enough to make possible the good's provision. (ii) Laboratory auction results which show that bidding in an auction process only reaches full WTP after a series of iterative learning periods (pp. 83, 89ff).

Regarding (i), we agree that the initial payment card amount is likely to be a low estimate because people may not initially fully face up to the "all or nothing" character of the situation. This raises the question of how to capture the understated consumer surplus. We do not think the followup bidding game is necessarily the answer. The procedure of bidding the price up in the "would you pay \$1 more" manner, runs a strong risk of twisting people's arms to go higher than they would really want to go (or would vote for in a referendum). This is because the followup bidding game procedure places people in the awkward social position of having to say no to the inferred request of the interviewer that they increase their amount by a mere \$1 (or whatever the interval is) for a socially desirable public good. One way to iterate with less chance of implied value type biases would be to say something to the effect that "if your amount was not enough to accomplish the change and it would have to be foregone unless more money was available, would you pay anything more?" If the person answered yes, he or she could then be asked to say the maximum additional amount they would pay before they would prefer to forgo the change.

Regarding (ii), we agree that CV scenarios should include iterative elements which permit learning to take place. And the more unfamiliar the good, the greater the need for these elements. We disagree with the notion that using a bidding game or multiple administrations of the instrument in a panel design are necessary to accomplish this end. In the first place, the evidence cited in the draft report that practice round (s) are needed in experiments which use Vickery auctions is not persuasive because a second price auction is not an intuitive institution for many people. Likewise, the data presented on p. 95f suggests that a WTA format is also not intuitive. However, unlike either of these formats, the WTP format appears to be simpler and more understandable. Second, as we will argue at more length below, use of a referenda model instead of a marketd goods model suggests that iteration of the kind proposed in the report is not an imperative design feature for CV surveys.

Thus it does not appear to us to be the case that an extended period of time or numerous iterations of a bidding game format which uses the WTP format are necessary to arrive at the true value. The data presented in the report's Figure 4.1 appear to support our contention. In this figure, the experimental iterations made a minor difference at best in the WTP bid compared with very large differences in the WTA bid. We firmly believe, however, that respondents in CV surveys do need to get into the game, and that scenarios should make every effort to help them to realize how it works. One technique we have found to be helpful (Carson and Mitchell 1984) is to provide respondents with opportunities to reconsider their answers at various points in the course of the questionnaire.

3. Payment Card. The report says relatively little about the payment card elicitation procedure except to describe some of the

experiments which have compared the use of payment cards with several other elicitation methods. It is important to emphasize that while payment cards formats were designed to avoid starting point bias, payment cards are not immune to other forms of bias involving implied value cues. Because of this, decisions about the number of dollar amounts which are displayed in a given card, their range, the size of the increments, and (if used) the nature of the anchors, must take into account the nature of the good and the expected value range. If the appropriate decisions are made, payment card bias can be minimized. If inappropriate decisions are made, the potential for bias is considerable. To take an extreme example, the use of a payment card whose first two numbers are \$0 and \$25 could lead to a substantial upward bias when valuing a good whose expected value (perhaps determined by in-depth pretests) is in the range of 12 - 15 dollars. Even when respondents are instructed, as they should be, to pick any number in between the amounts shown on the card, in our experience they tend to limit the choice to the numbers on the card. As a result, respondents who have a true value of \$15 for the good may be influenced by the design of this payment card to pick the \$25 amount and therefore overstate their WTP amount.

Not enough is known about the effect of changing the various parameters of payment cards and more research is called for. Research which tests the influence of extreme differences in the upper bound is not very informative, however, since different mean WTP amounts are to be expected under this condition. The most useful research would focus on the effect of parameters within the range of reasonable values such as the effect of upper bounds at 3, 5, and 7 times the expected average value.

B. THE AUCTION ANALOGY (Chapter IV.)

This chapter develops an auction analogy for CV studies which is based on auctions for private goods, where the true price can be established. Since most CV surveys value public not private goods, the relevance of auction theory to the provision of collective goods needs to be established, something the draft version of the report does not do.

1. Second Price Auction. The draft report recommends the second price auction model for CV surveys. While a second price auction has desirable properties, it is impossible to simulate in a CV study without greatly complicating the scenario. And the use of increments in an English auction, if they are large relative to the price, make it no longer equivalent to a second price auction (Carson and Foster, 1984), thus qualifying conclusion 6 on page 88 of the draft report.

Putting aside for the moment the collective properties of public goods, CV surveys might be viewed as analogous to first price auctions in the sense that the respondent, like the bidder in such an auction, believes he or she will have to pay the price if the good is provided.⁵ In CV surveys, such a belief has the desirable property that if it does induce a bias, it is to underestimate the WTP for the good since first price auctions yield prices at or below that of second price auctions. Any difference between a first and a second price (if such a thing could be obtained for a good valued in a CV survey) is likely to be caused either by strategic behavior or by the respondent's undervaluing the good because of not having faced up to the implications of not receiving it. In both cases, the scenario can help overcome these problems by emphasizing the potential for everyone being excluded from the good if it is not provided.

C. UTILITY OF LABORATORY EXPERIMENTS

The fact that CV studies value public goods raises serious questions about the use of laboratory experiments to determine optimal CV scenario formats such as those advocated in Chapter IV. In the absence of a good which can really be sold through a second price auction, what criterion will such experiments use in order to evaluate various alternative design features? On the other hand, if the experiments use a good which can be bought and sold (such as Bishop and Heberlein's hunting permits), the direct applicability of these findings based on a private good to situations with public goods is uncertain. Also relevant to the utility of experiments is the fact that CV surveys normally value goods by interviewing fairly large, random samples. The notion, which the report advocates, of applying a tatonnement voting process, which requires unanimity, to any but a very small group seems highly impractical to us. Quite apart from its impracticality, we fail to understand why unanimity is necessary since the likelihood of strategic behavior in properly designed CV studies has been shown to be acceptably low.

In our view, what is needed are not experiments aimed at developing mechanisms to simulate second price auctions, which are likely to be unsuitable to the field conditions faced by CV studies, but laboratory and field work which illuminate the conditions under which certain biases occur in the field and which give us greater understanding of what goes on in people's minds when they answer WTP questions. Desvousges and Smith's use of focus groups is a case in point, as is their work in debriefing interviewers to better understand the responses to their Monongahela survey. Much more work needs to be done on this count. For example, we need to know how people tend to partition environmental goods in their minds in order to better understand the part-whole problem identified by Randall and Hoehn. In-depth interviewing of a few respondents or the debriefing of participants in a relevant experiment can potentially yield insights on this topic which could really make a difference in field applications.

D. THE EFFECT OF HYPOTHETICALITY (Chapter V)

1. Hypothetical bias. We believe the use of this term is confusing. Although the hypothetical character of CV studies has several potential effects, one of which is to increase the likelihood of certain other types of bias (the other is to increase the random quality of the answers), there is no unique hypothetical bias.^{6/}

2. Tests of Whether Actual vs. Hypothetical Payment Makes a Difference. Our review of the literature leads us to question the draft report's conclusions: (i) "The literature abounds with evidence that suggests that ... actual vs. hypothetical payment does result in different choices (p. 107, emphasis in the original) and (ii) ... the quality of empirical measures of value from the HPM per se are far from a level where they might be regarded as accurate, in some sense, estimates for market values attributable to public goods (p. 110)."

The evidence, at least as we read it, is much more equivocal on both of these points. We begin with (i) above. The draft report cites four bodies of evidence in support of this contention, several of which do not support it and others of which support it much less than suggested. The first is Bohm's work whose conclusion that people will act "irresponsibly" where no payment is involved you accept as proved. In our 1981 report, we reanalyzed Bohm's data and showed that this conclusion rested on one outlier. More recently (Mitchell and Carson, 1984), we have further reanalyzed Bohm's findings in light of recent experimental work (Marwell and Ames etc.). Quite apart from our original criticism, which still holds, we now view Bohm's treatment VIb (which is essentially a first price auction where the top ten out of 100 bids were accepted) as representing the closest approximation to the true WTP of all of his treatments. In light of this, it is significant that the mean bid for this treatment, K10.3, is almost identical to the mean bid for VIa, the only hypothetical treatment in his experiment.^{7/} The second body of evidence is Bishop and Heberlein's original goose hunting study. In this case the draft report accepts our criticism of Bishop and Heberlein's finding. Presuming that our critique of these two important studies is correct, this leaves us with only two pieces of evidence for the draft report's finding that the literature "abounds" with evidence that actual and hypothetical payment result in different choices.

The third study is Coursey et al., 1983, an unpublished experiment. Our reading of this paper suggests that the difference found in this study has to do with WTP vs. WTA and not with hypothetical WTP vs. real WTP. Since the WTA/WTP issue has its own complexity (e.g. Michael Hanneman has shown that the Willig bounds do not hold for discrete choice situations) and since WTP is the format used in most CV studies, this study is largely irrelevant to the generalization.

The fourth body of evidence are "tests of actual vs. hypothetical payment on decision strategies reported in the psychological literature." You cite Paul Slovic's 1969 conclusion that real vs. hypothetical gains or losses made a difference in people's decision strategies, as "typical" of the findings of these studies (p. 108). Because our understanding of this literature was that it also contained a number of studies, such as Grether and Plott's, which tested certain findings (such as preference reversals) under both conditions with the opposite conclusion, we called Paul Slovic to see what studies we had missed. In our conversation he made the following points: 1) Generally speaking, the literature on this topic shows equivocal findings. 2) Very few studies have examined the effects of hypothetical vs. real payments directly as his 1969 study did. His study was very sensitive to decision strategy in that it looked at gambles. 3) There are a lot of studies similar to Grether and Plott's which find that observed effects hold under both conditions. In the absence of other evidence, we conclude that the matter is less clear than the draft report's presentation would suggest and that Slovic's 1969 study doesn't really support the pessimistic conclusion.

Thus the evidence for actual vs. hypothetical payments making a difference is very weak. What about the other side of the question? According to the report (p. 108) "there is little if any evidence that would support the hypothesis that actual payment = hypothetical payment." It is true that there is very little direct evidence for this hypothesis, just as there is little direct evidence for the reverse hypothesis. Studies which attempt to predict behavior on the basis of measures of behavioral intentions provide some useful indirect evidence on this issue, however. You cite one such study -- Kogan and Wallach (1968); there are a number of others in the attitude-behavior literature which bear on this question. There is also some relevant work in the market research literature on "concept testing" (Moore 1982). These studies demonstrate that, under certain conditions, surveys can have excellent predictive value.^{8/}

To summarize, we argue the following: (1) By no means does the literature abound with evidence that actual vs. hypothetical payment results in different choices. The evidence, we find, is very weak on this point. (2) Although there is little direct evidence for the opposite hypothesis, important indirect evidence is available. (3) The essential fact is that the literature simply does not permit an authoritative statement to be made one way or the other.

In making this argument we do not mean to imply that the hypothetical character of CV studies is unproblematic. Indeed, we believe the greatest methodological problems with the CV method stem from their hypothetical character. Nevertheless, there are reasonable grounds in the literature to support the idea that carefully designed hypothetical payment situations can approximate actual payment situations with sufficient accuracy to be a useful component of benefit/cost analysis.

We now turn to the draft report's second finding that the quality of hypothetical CV values "are far from a level where they might be regarded as

accurate." Presuming that this means that even well designed CV studies with good samples and excellent field work are inevitably very inaccurate, this strong statement is simply not supported by the evidence provided in this chapter. Nor do we believe the statement captures the reality of what the past decade of research on CV has found. To repeat, our own view is that while it is very difficult to obtain unbiased or minimally biased CV estimates, properly designed CV studies are possible and they can obtain benefit measures with acceptable levels of accuracy.

E. POLITICAL MARKETS AND REFERENCE OPERATING CONDITIONS (Chapter VI.)

The draft report argues that accuracy in CV studies can only be measured by the degree to which these studies replicate what is termed the "key reference operating conditions for the market institution." "Market institution" is defined as markets in private goods where goods are traded frequently through a process of competing bids and offers. In such a market, according to the report, people gain information through the process of frequent purchase. The requirements imposed by the market model then determines the first two reference operating conditions (ROC's).

An alternative framework is suggested by the large body of theory developed by economists and political scientists (e.g. Enelow and Hinich, 1984; Bergstrom, Rubinfeld and Shapiro, 1982; Deacon and Shapiro, 1975) on political markets which, after all, are how public goods are supplied. Here the particular form of the political market most relevant to CV is the referendum. In a referendum, a voter is faced with a one-time (or at best with a very infrequent) choice on a predetermined policy package to which they must vote yes or no before the outcome of the referendum is known. If the particular issue comes up in a subsequent referendum, it is likely to pose a fairly sizable change in the policy package.

The referendum framework suggests a somewhat different set of reference operating conditions. ROC's from referenda and the psychological research which point to distortions in decision processes appear to us to consist Of the following:

- 1) Respondents must understand the commodity to be valued, how it will be provided and how it will be paid for,
- 2) They should be given as simple a choice as possible.
- 3) There must be little uncertainty about the provision of the good.
- 4) WTP, not WTA, measures are elicited.
- 5) Outliers should not be permitted unduly to influence the results.
- 6) Respondents should be permitted to abstain from the valuation process.

Items 3) and 4) are identical to those derived from the market model and presented on page 199 of the draft report. The other items bear some explanation.

Item 1): For a referendum to measure people's true WTP for the commodity, the voters should understand the nature of the commodity, its method of provision and the consequences of its implementation. (In practice, some people make uninformed decisions in referenda just as they do in the marketplace.) In CV studies, the scenario must be able to accurately convey this information to respondents with widely varying educational attainments and life experience. Understanding is usually made easier if the respondent has had experience with the good. But it is worth noting that experience is not necessarily an advantage since familiarity can interfere with understanding by leading respondents to jump to mistaken conclusions about the scenario's elements. For example, the use of a park entrance fee as a payment vehicle for valuing park amenities is something

which many respondents have experienced. Asking for a maximum WTP amount by use of such a vehicle faces the problem that respondents are likely to hold a conception of a "fair" or normative park entrance fee based on their experience. Thus, while they may in fact have a true WTP amount of \$13 for a particular amenity, they may bid less than this amount if their view of a maximum fair entrance fee is \$10 or \$7 or \$5 etc.

Item 2): Referenda pose issues in terms of a yes/no decision for a particular level of provision of a good at a given price. CV scenarios should strive for as simple a choice as possible within the methodological limitations of survey research and modest sample sizes. The potential for yea-saying bias limits the application of a direct imitation of the referenda format as do the large sample sizes required by formats using dichotomous answers to priced levels of the good. Where the choice is complex, respondents should be provided opportunities to change their decisions after they have gone through the valuation process and understand its full implications. Note that referenda are often one-time exercises where voters vote on items about which they may not have had prior valuation and choice experience (e.g. nuclear referenda, water bond issues etc.).

Item 5): Referenda use a majority or 2/3 rule for deciding whether or not a public commodity is to be supplied. In either case a small minority (ie, outliers) do not determine the decision made.

Item 6): Participation in referenda is voluntary. Voters can choose whether or not to go to the polls and once there, they can choose whether or not to vote on particular issues. CV studies should not "require" answers to the WTP questions from respondents who would prefer not to answer because they are not interested in the valuation exercise, are confused by it, can't determine what value they hold for the good etc, If they do, the quality of the data will suffer. However, in order to obtain a valid population estimate, the effect of nonresponses must be adjusted for by use of Scientific sampling and missing data estimation techniques.

F. ON THE QUESTION OF ACCURACY (Chanter VI)

The report's argument for a fifty percent error range is not convincing. The size of the range appears to be arbitrarily chosen and the statistical properties of the range are not well defined. Is the range, for example, meant to include the sampling variance? We do not think it should. If it doesn't, and sampling variance is added to the fifty percent error range, studies with small samples will be expected to produce very large estimate ranges.

Not all CV studies in the literature produce findings which are equally valid. Some suffer from severe methodological problems which bias the results. Some have very small sample sizes which affect the statistical tests of differences. To talk about the general accuracy of CV studies in terms of an arbitrarily chosen and imprecisely defined $\pm 50\%$ criterion, ignores this problem and seems to suggest that as long as a study meets the ROC's specified in the report, it will provide a reasonably satisfactory "rough" estimate. Quite apart from our views about whether the ROC's recommended in the draft report are the most appropriate ones, both the report's and ours are too general to be of much help in providing criteria by which a CV study can be evaluated. The key questions are: How does one tell a "good" study in the sense of a properly conducted CV study, from a "poor" one? What improvements are needed to increase the accuracy of CV studies? Which improvements promise the most payoff? These are the kinds of issues which could have been explored to advantage in the report's discussion of accuracy.9/

ENDNOTES

Appendix To Chapter XIII

- *) The State of the Arts Assessment of the Contingent Valuation Method Draft Report and the conference on this report represent an important milestone in the development of the contingent valuation method. They addressed a number of the important and difficult issues associated with the methodology, some of which were overdue for attention and others of which have engaged the thought of CV practitioners for some time. These remarks were prepared before the Palo Alto Conference in response to an invitation from the authors for critical comments. We appreciate the opportunity to participate in the debate raised by this stimulating report. What follows are our views on a set of issues where we disagree with the report's presentation. We have revised these remarks somewhat since their original formulation. All page numbers refer to the original draft report.
- 1) Studies such as Walsh and Gilliam (1982), which are cited in the report, appear to be irrelevant to the issue, at least as described.
 - 2) See Desvousges et al., 1983, pp. A-1ff for a good discussion of this topic, but note that the coefficients of variation for given CV studies in Table A-2 are incorrectly estimated and are much too small because they calculated the coefficient of variation with the standard error of the mean instead of the standard deviation.
 - 3) A pretest which we ran in the summer of 1983 illustrates the potential consequence of small sample sizes for hypothesis testing. In addition to pretesting our water quality instrument, we wanted to test the effect of using payment Cards with and without anchors. Our usable N's were 37 and 39 for the two treatments and the coefficient of variations were roughly 2.0, a size similar to that found in many CV studies. If we wished to use standard comparison of means tests to detect a 25 percent or greater difference between the two treatments, with (i) a ten percent chance of rejecting the hypothesis of no difference where there is a difference and (ii) a five percent chance of accepting a difference only 5 percent of the time when in fact no difference is present, we would have had to have a sample size of approximately 2000 for each treatment. Expressed another way, given our actual sample size, the mean of the second treatment would have had to have been 75 percent larger (or smaller) than that of the first treatment before we could have rejected the null hypothesis. Because income is a good predictor of the WTP amounts in this case, we were able to assume a log normal distribution which enabled us to use a powerful test of the hypothesis that the medians of the original distribution (mean of the log distribution) were significantly different. (They were not.)

In cases where income is not a good predictor, however, such a procedure is not possible and with small samples the deck is potentially stacked against finding a difference.

- 4) Comparisons of the mean bids for nonusers showed the \$125 starting point bids were almost double those for the \$25 starting point treatment.
- 5) The CV interpretation of a second price auction is that the respondent bids believing he or she will have to pay the average WTP amount if the good is provided. This situation has no known desirable properties.
- 6) We develop this argument (and some of the others which we mention in this critique) more fully in Mitchell and Carson (1984).
- 7) Note that subjects participating in VIb had the most "iterative" experience of any of Bohm's subjects, as the same sample also took part in VIa.
- 8) This assertion is based on an analysis presented in Mitchell and Carson (1984).
- 9) The draft report does not address these issues and often ignores their implications. For example, Table 6.12 presents the mean values obtained by studies which compare CV and indirect market methods of valuation. Lacking from this table are the studies' sample sizes and standard deviations which, would indicate a) whether the differences are significant and b) whether the imprecision is due to the CV study, the indirect study or to both methods. Likewise, the issue of sampling is not discussed in the report despite its implications for accurate benefit estimates from CV data. Many CV studies in the literature provide no information or very scanty information about the sampling plan and its execution. Errors in aggregation based on faulty sampling could easily be in the 50 percent range. Another type of aggregation problem which the study does not discuss is the sometimes high item-nonresponse rate in CV studies. A greater number of respondents in some CV surveys based on random samples fail to answer the WTP questions than fail to answer questions in ordinary surveys. Within limits, this is understandable (WTP questions are demanding) and desirable (better to have don't knows than guesses). In order to derive accurate population estimates from such data, however, the use of approximation techniques is required. Our preliminary work on this topic suggests possible errors due to this factor alone in the 10 - 25 percent range.

REFERENCES

- Ajzen, I. and M. Fishbein, 1977. "Attitude-Behaviors Relations: A Theoretical Analysis and Review of Empirical Research," Psychological Bulletin, 84, pp. 888-918.
- Akerlof, G.A. and W.T. Dickens, 1982. "The Economic Consequences of Cognitive Dissonance," American Economic Review, 72(3).
- Arrow, K.J., 1982. "Risk Perceptions in Psychology and Economics," Economic Inquiry, 20, pp. 1-9.
- Arrow, K.J., 1951. Social Choice and Individual Values, New York, Wiley.
- Banford, N.D., J.I. Knetsch and G.A. Mauser, 1977. "Compensating and Equivalent Variation Measures of Consumer's Surplus: Further Survey Results," unpublished paper, Dept. of Economics and Commerce, Simon Fraser University, Vancouver, B.C., Canada.
- Bartik, T.J. and V.K. Smith, 1984. "Urban Amenities and Public Policy," Vanderbilt University Working Paper No. 84-W18, May.
- Batie, S.S. and L. Shabman, 1979. "Valuing Non-Market Goods - Conceptual and Empirical Issues: Discussion," American Journal of Agricultural Economics, 61(5), pp. 931-32.
- Belovicz, M.W., 1979. "Sealed-Bid Auctions: -Experimental Results and Applications," in V.L. Smith (ed.), Research in Experimental Economics, Vol. 1, Connecticut, JAI Press, Inc.
- Bergson, A., 1938. "A Reformulation of Certain Aspects of Welfare Economics," Quarterly Journal of Economics, 68(2).
- Bergstrom, T.C., D.L. Rubinfeld and P. Shapiro, 1982. "Micro-Based Estimates of Demand Functions for Local School Expenditures," Econometrica, 50(5), pp. 1183-1205.
- Bishop, R.C., 1982. "Option Value: An Exposition and Extension." Land Economics, 58, pp. 1-15.
- Bishop, R.C., T.A. Heberlein and M.J. Kealy, 1983. "Contingent Valuation of Environmental Assets: Comparison With a Simulated Market," Natural Resources Journal, 23(3), pp. 619-34.
- Bishop, R.C. and T.A. Heberlein, 1984. "Contingent Valuation Methods and Ecosystem Damages from Acid Rain," Agricultural Economics Staff Paper Series, No. 217, Department of Agricultural Economics, University of Wisconsin - Madison.
- Blank, F.M., D.S. Brookshire. T.D. Crocker, R.C. d'Arge, R.L. Horst and R.D. Rowe, 1978. Valuation of Aesthetic Preferences: A Case Study of the Economic Value of Visibility, prepared for the Electric Power Research Institute, Inc. (EPRI), Palo Alto, CA, Contract #RP7855-2,

October 6.

- Blumburg, G.J., 1984. "Application and Assessment of the Contingent Valuation Method for Federal Hazardous Waste Policies in the Washington, D.C. Area." Unpublished Ph.D dissertation, University of New Mexico, Albuquerque.
- Bohm, P., 1971. "An Approach to the Problem of Estimating Demand for Public Goods," Swedish Journal of Economics, 73(1), pp. 55-66.
- Bohm, P., 1972. "Estimating Demand for Public Goods: An Experiment," European Economic Review, 3(2), pp. 111-130.
- Boulding, K.E., 1975. The Image, Ann Arbor, University of Michigan Press.
- Boyle, K.J. and R.C. Bishop, 1984. "Lower Wisconsin River Recreation: Economic Impacts and Scenic Values,." Staff Paper No. 216, Department of Agricultural Economics, University of Wisconsin - Madison.
- Boyle, K.J., R.C. Bishop and M.P. Walsh, forthcoming, "Starting Point Bias in Contingent Valuation Bidding Games," Land Economics.
- Bradford, D.F., 1970. "Benefit-Cost Analysis and Demand Curves for Public Goods," Kyklos, 23, PP. 775-791.
- Brookshire, D.S. and T.D. Crocker, 1981. "The Advantages of Contingent Valuation Methods for Benefit-Cost Analysis," Public Choice, 36(2), pp. 235-252.
- Brookshire, D.S. and R.C. d'Arge, 1979. "Resource Impacted Communities: Economics, Planning and Management," paper prepared for the 4th U.S. - U.S.S.R. Symposium on Comprehensive Analysis of the Environment, Jackson, Wyoming, October 21-27, pp. 1-55.
- Brookshire, D.S., R.C. d'Arge, W.D. Schulze and M.A. Thayer, 1981. "Experiments in Valuing Public Goods," Advances in Applied Microeconomics, Vol. 1, V.K. Smith, ed., Connecticut, JAI Press, Inc., pp. 123-172.
- Brookshire, D.S. and L.S. Eubanks, 1978. "Contingent Valuation and Revealing the actual Demand for Public Environmental Commodities," in Public Choice in New Orleans, G. Tullock (ed.), selected paper of the Public Choice Society, New Orleans.
- Brookshire, D.S., L.S. Eubanks and A. Randall, 1983a, "Estimating Option Prices and Existence Values for Wildlife Resources," Land Economics, 59, pp. 1-15.
- Brookshire, D.S., B.C. Ives and W.D. Schulze, 1976. "The Valuation of Aesthetic Preferences," Journal of Environmental Economics and Management, 3, pp. 325-46.

- Brookshire, D.S., A. Randall, R.C. d'Arge, L.S. Eubanks, J.R. Stoll, T.D. Crocker and S. Johnson, 1977. Methodological Experiments in Valuing Wildlife Resources: Phase I Interim Report to the United States Fish and Wildlife Service, USFWS Grants #14016-00009-77-002 and 14016-00009-77-003, November.
- Brookshire, D.S., A. Randall, and J.R. Stoll, 1980. "Valuing Increments and Decrements in Natural Resource Service Flows," American Journal of Agricultural Economics, 62(3), pp. 478-488.
- Brookshire, D.S., W.D. Schulze, M.A. Thayer and R.C. d'Arge, 1982a, "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches," American Economic Review, 72(1), pp. 165-177.
- Brookshire, D.S., W.D. Schulze and R. Whitworth, 1983b. An Economic Analysis of the Private Sector Benefits and Potential for Cost Recovery of the National Coal Resources Data Systems, report to the U.S. Geological Survey, Reston, VA, July.
- Brookshire, D.S., M.A. Thayer, J. Tischirhart, and W.D. Schulze, 1984. "A Test of the Expected Utility Model: Evidence from Earthquake Risks," manuscript, University of Wyoming, Laramie, forthcoming in Journal of Political Economy.
- Brown, J. and S. Rosen, 1982. "On Estimation of Structural Hedonic Price Models," Econometrica, 50(3), pp. 765-68,
- Brown, W., A. Singh, and E. Castle, 1964. An Economic Evaluation of the Oregon Salmon and Steelhead Sports Fishery, Oregon Agricultural Experiment Station Bulletin, No. 74, Corvallis.
- Buchanan, J. and G. Tullock, 1962. The Calculus of Consent, Ann Arbor, University of Michigan Press.
- Burness, H.S., R.G. Cummings, A.F. Mehr and M.S. Walbert, 1983. "Valuing Policies Which Reduce Environmental Risk," Natural Resources Journal, 23(3), pp. 675-682.
- Carson, R.T. and W.E. Foster, 1984. "A Theory of Auctions from the Auctioneer's Perspective," presented at the Econometric Society Meetings, Stanford University.
- Carson, R.T. and R.C. Mitchell, 1984. "A Reestimation of Bishop and Heberlein's Simulated Market-Hypothetical Markets-Travel Cost Results Under an Alternative Assumption," Discussion Paper No. D-107 Washington, D.C., Resources for the Future, forthcoming, American Journal of Agricultural Economics.
- Carson, R.T. and R.C. Mitchell, 1984, "National Fresh Water Quality Benefits: Findings from a New National Contingent Valuation Study" presented at the Association of Environmental and Resource Economists Meeting at Cornell University, August. Discussion Paper No. D-124, Washington, D.C., Resources for the Future.

- Carson, R.T. and R.C. Mitchell, 1984. "Non-Sampling Errors in Contingent Valuation Surveys," Discussion Paper D-120, April, Washington, D.C., Resources for the Future.
- Cassady, R., 1967. Auctions and Auctioneering, Berkeley, University of California Press.
- Cesario, F.J., 1976. "Valuing Time in Recreation Studies," Land Economics, p. 52.
- Chew, Soo-Hung and K.R. Macgrimmon, 1979. "Alpha Utility Theory, Lottery Composition and the Allais Paradox," working paper, Faculty of Commerce and Business Administration, University of British Columbia, September.
- Ciriacy-Wantrup, S.V., 1952. Resource Conservation: Economics and Policies, Berkeley, University of California Press.
- Clarke, E.H., 1971. "Multipart Pricing of Public Goods," Public Choice, 11, PP. 17-33.
- Clawson, M., 1959. "Methods of Measuring the Demand for Outdoor Recreation," reprint No. 10, Washington, D.C., Resources for the Future.
- Coppinger, V.M., V.L. Smith and J.A. Titus, 1980. "Incentives and Behavior in English, Dutch and Sealed-Bid Auctions," Economic Inquiry, 18(1), pp. 1-22.
- Coursey, D.L. and V.L. Smith, 1982. "Experimental Tests of an Allocation Mechanism for Private, Public or Externality Goods," mimeo, Department of Economics, University of Arizona.
- Coursey, D.L. and H. Nyquist, 1983. "Application of Robust Estimation Techniques in Demand Analysis," manuscript, University of Umea, Sweden, October, submitted to Review of Economics and Statistics.
- Coursey, D.L., W. D. Schulze and J. Hovis, 1983. "A Comparison of Alternative Valuation Mechanisms For Non-Market Commodities," unpublished manuscript, Department of Economics, University of Wyoming, December.
- Cox, J.L., B. Roberson and V.L. Smith, 1982. "Theory and Behavior of Single Price Auctions," in V.L. Smith (ed.), Research in Experimental Economics, Vol. 2, Connecticut, J.A.I. Press, Inc.
- Crandall, R.W., 1983. Controlling Industrial Pollution: The Economics and Politics of Clean Air. Washington, D.C., The Brookings Institution.
- Crocker, T.D., 1984. "On The Value of the Condition of a Forest Stock," unpublished manuscript, Department of Economics, University of Wyoming.
- Crocker, T.D. and R.G. Cummings, 1984. "On Valuing Deposition-Induced Materials Damages: A Methodological Inquiry," in D.D. Adams and

- W. Page (eds.), Acid Deposition: Environmental and Economic Impacts, New York, Plenum Press, forthcoming.
- Crocker, T.D. and B.A. Forster, 1984. "Nonconvexities in the Control of Acid Deposition: Still a Problem," Journal of the Air Pollution Control Association, forthcoming.
- Cronin, F.J., and K. Herzeg, 1982. Valuing Nonmarket Goods Through Contingent Markets, Pacific Northwest Laboratory, Richland, Washington, PNL-4255.
- Cummings, R.G., H.S. Burness and R.D. Norton, 1981. "Measuring Household Soiling Damages from Suspended Air Particulates: A Methodological Inquiry," Report to the U.S. Environmental Protection Agency, R805059010, Washington, D.C., April.
- Cummings, R.G., W.D. Schulze, D.S. Brookshire and S.D. Gierking, 1982. "Measuring the Elasticity of Substitution of Wages for Municipal Infrastructure: A Comparison of the Survey and Wage Hedonic Approach," manuscript, Department of Economics, University of New Mexico, Albuquerque.
- Cummings, R.G., W.D. Schulze and A.F. Mehr, 1978. "Optimal Municipal Investment in Boomtowns: An Empirical Analysis," Journal of Environmental Economics and Management, 5.
- Dasgupta, A.K. and D.W. Pearce, 1978. Cost-Benefit Analysis: Theory and Practice, New York, MacMillan.
- Daubert, J. and R. Young, 1981. "Recreational Demands for Maintaining Instream Flows: A Contingent Valuation Approach", American Journal of Agricultural Economics, 63(4), pp. 666-676.
- Davis, R.K., 1963a. "Recreational Planning as An Economic Problem," Natural Resource Journal, pp. 238-249.
- Davis, R.K. 1963b. "The Value of Outdoor Recreation: An Economic Study of the Maine Woods", unpublished Ph.D. dissertation, Harvard University.
- Deacon, R. and P. Shapiro, 1975. "Private Preference for Collective Goods Revealed Through Voting on Referenda," American Economic Review, 65(5), pp. 943-955.
- Desvousges, W.H. and V.K. Smith, 1982. The Basis for Measuring the Benefits of Hazardous Waste Disposal Regulation, Raleigh, Research Triangle Institute.
- Desvousges, W.H., V.K. Smith, and M.P. McGivney, 1983. A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements, EPA-230-05-83-001, March.

- Devletoglou, N.E., 1971. "Thresholds and Transaction Costs," Quarterly Journal of Economics, 85(1).
- Dubey, P. and M. Shubik, 1980. "A Strategic Market Game with Price and Quantity Strategies," Zeitschrift fur Nationalokonomie, 40.
- Enelow, James M. and Melvin J. Hinich, 1984. The Spatial Theory of Voting: An Introduction. Cambridge, Cambridge University Press.
- Engelbrecht-Wiggans, R., 1980. "Auctions and Bidding Models: A Survey," Management Science, 26.
- Evans, John S., 1984. "Theoretically Optimal Environmental Metrics and Their Surrogates," Journal of Environmental Economics and Management, 71, pp. 18-27.
- Feather, T. 1959. "Subjective Probability and Decisions Under Uncertainty," Psychology Review, 66, pp. 150-64.
- Feenberg, D. and E.S. Mills, 1980. Measuring the Benefits of Water Pollution Abatement, New York, Academic Press.
- Ferejohn, J.A. and R.G. Noll, 1976. "An Experimental Market For Public Goods: The PBS Station Program Cooperative," American Economic Review Proceedings, 66, pp. 267-273.
- Ferejohn, J.A., R. Forsythe and R. Noll, 1979a. "An Experimental Analysis of Decision Making Procedures for Discrete Public Goods: A Case Study of 2 Problem of Institutional Design," in V.L. Smith (ed.), Research in Experimental Economics, Vol. 1, Connecticut, JAI Press, Inc.
- Ferejohn, J.A., R. Forsythe and R. Noll, 1979b. "Practical Aspects of the Construction of Decentralized Decision Making Systems for Public Goods," in C. Russell (ed.), Collective Decision Making: Applications from Public Choice Theory, Baltimore, Johns Hopkins University Press.
- Ferejohn, J.A., R. Forsythe, R. Noll and T.R. Palfrey, 1982. "An Experimental Examination of Auction Mechanisms for Discrete Public Goods," in V.L. Smith (ed.), Research In Experimental Economics, Vol. 2, Connecticut, JAI Press, Inc.
- Fischhoff, B., 1975. "Hindsight Is Not Equal To Foresight: The Effect of Outcome Knowledge On Judgement Under Uncertainty," Journal Of Experimental Psychology, 104(1), pp. 288-99.
- Fischhoff, B., P. Slovic and S. Lichtenstein, 1982. "Lay Foibles and Expert Fables in Judgements About Risk," The American Statistician, 36(2), pp. 240-255.
- Fischhoff, B., 1982. "For Those Condemned to Study the Past: Heuristics and Biases in Hindsight," in Judgements Under Uncertainty: Heuristics and Biases, in D. Kahneman, P. Slovic and A. Tversky (eds.), Cambridge, Cambridge University Press, pp. 335-354.

- Fishbein, M., and I. Ajzen, 1975. "Attitudes Toward Objects as Predictors of Single and Multiple Behavioral Criteria," Psychology Review, 81, pp. 59-74.
- Fisher, W.L., 1984. "Measuring the Economic Value of Fishing and Hunting: A Conceptual Overview," Paper presented at the Annual Meeting of Omicron Delta Epsilon, Lycoming College, Williamsport, Pa., April 16.
- Flinn, J. and J. Heckman, 1983, "Are Unemployment and Out of the Labor Force Behaviorally Distinct States?" Journal of Labor Research 1, pp. 28-43.
- Forsythe, R. and R.M. Isaac, 1982. "Demand-Revealing Mechanisms for Private Goods Auctions," in V.L. Smith (ed.), Research In Experimental Economics, Vol. 2, Connecticut, JAI Press, Inc.
- Freeman, A.M., 1979a. Benefits of Environmental Improvement, Baltimore, Johns Hopkins University Press.
- Freeman, A.M., 1979b. "Approaches To Measuring Public Goods Demands," American Journal of Agricultural Economics, 61(5) pp. 975-20.
- Freeman, A.M., 1979c. "Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues," Scandinavian Journal of Economics, 81(2), pp. 154-173,
- Freeman, A.M., 1984. "The Sign and Size of Option Value," Land Economics, 60, pp. 1-13.
- Freeman, B. and L. Medoff, 1982. "(Why Does the Rate of Youth Labor Force Activity Differ Across Surveys," in R.B. Freeman and D.A. Wise (eds.), The Youth Labor Market Problems: Its Nature, Causes, and Consequences. Chicago, University of Chicago Press.
- Friedman, M. 1953. Essays In Positive Economics, Chicago.
- Gallagher, D.R and V.K. Smith, 1984. "Measuring Values for Environmental Resources Under Uncertainty," forthcoming Journal of Environmental Economics and Management.
- Georgescu-Roegen, N. 1958. "Threshold in Choice and The Theory of Demand," Econometrica, 26, pp. 157-68.
- Gerking, S. and W.D. Schulze, 1981. "What Do We Know About Benefits of Reduced Mortality From Air Pollution Control?" American Economic Review, 71(2), pp. 228-40.
- Gordon, I.M. and J.L. Knetsch, 1979. "Consumer's Surplus Measures and the Evaluation of Resources," Land Economics, 55, pp. 1-10.
- Graham, D.A., 1981. "Cost-Benefit Analysis Under Uncertainty," American

- Economic Review, 71, pp. 715-725.
- Gramlich, F. 1977. "The Demand for Clean Water: The Case of the Charles River," National Tax Journal, 30(2), pp. 183-194.
- Green, M.W. 1942. "Sucrose Octa-Acetate as a Possible Bitter Stomachic," Bulletin of the National Formulary, Pharmaceutical Association, 10, pp. 131-133.
- Greenley, D.A., R. Walsh and R.A. Young, 1981. "Option Value: Empirical Evidence From A Case Study of Recreation and Water Quality," Quarterly Journal of Economics, pp. 657-673.
- Greenley, D.A., R.G. Walsh and R.A. Young. 1982. Economic Benefits of Improved Water Quality, Boulder, Westview Press, pp. 164.
- Gregory, R.S., 1982. "Valuing Non-Market Goods: An Analysis of Alternative Approaches," unpublished Ph.D. dissertation, University of British Columbia, Vancouver.
- Grether, D.M. and C.R. Plott, 1979. "Economic Theory of Choice and The Preference Reversal Phenomenon," American Economic Review, 69, pp. 623-38.
- Groves, T., 1973. "The Incentives In Teams", Econometrica, 41.
- Groves, T. and J. Ledyard, 1977. "Optimal Allocation of Public Goods: A Solution to the 'Free Rider Problem'," Econometrica, 45(4), pp. 763-809.
- Hammack, J. and G.M. Brown, Jr., 1974. Water Fowl and Wet Lands: Toward BioEconomic Analysis, Baltimore, Johns Hopkins University Press for Resources for the Future.
- Hanemann, W.M. 1983b. "Welfare Evaluation with Simulated and Hypothetical Market Data: Bishop and Heberlein Revisited," Giannini Foundation Working Paper, No. 276.
- Heberlein, T.A., 1975. "Social Norms and Environmental Quality," Paper Presented at the American Association for the Advancement of Science Annual Meetings. New York, January.
- Hershey, J.C. and P.J.H. Schoemaker, 1980. "Risk Taking and Problem Context in the Domain of Losses: An Expected Utility Analysis," Journal of Risk and Insurance, 47, pp. 111-32.
- Hey, J.D., 1983. "Whither Uncertainty," Economic Journal, (Supplement), pp. 130-39.
- Heyne, P. 1983. The Economic Way of Thinking, 4th ed., Chicago, Science Research Associates, Inc.
- Hoehn, J.P. and A. Randall, 1983. "Benefit Aggregation and

- Disaggregation Across Policy Components," University of Kentucky, unpublished paper.
- Hoehn, J.P. and A. Randall, 1983b. "Contingent Valuation is a Satisfactory Benefit-Cost Indicator," University of Kentucky, unpublished paper.
- Hoehn, J. P. and A. Randall, 1983. "Incentives and Performance in Contingent Valuation," University of Kentucky, unpublished paper.
- Hori, H., 1975. "Revealed Preference for Public Goods," American Economic Review, 65.
- Hueth, D. and E.J. Strong, 1984. "A Critical Review of the Travel Cost, Hedonic Travel Cost and Household Production Models for Measurement of Quality Changes in Recreational Experiences," American Journal of Agricultural Economics, 66(15).
- Hurwicz, L., 1973. "The Design of Mechanisms for Resource Allocation," American Economic Review Proceedings.
- Isaac, R.M., 1983. "Laboratory Experimental Economics as a Tool in Public Policy Analysis," mimeo, Dept. of Economics, University of Arizona.
- Jarecki, H., 1976. "Bullion Dealing, Commodity Exchange Trading, and the London Gold Fixing: Three Forms of Commodity Auctions," in Y. Amihud (ed.), Bidding and Auctioning for Procurement and Allocation, New York, New York University Press.
- Just, R.E., D.L. Hueth and A. Schmitz, 1982. Applied Welfare Economics and Public Policy, Englewood Cliffs, Prentice-Hall, Inc.
- Kahneman, D. and A. Tversky, 1972. "Subjective Probability: A Judgement of Representativeness," Cognitive Psychology, 3(3), pp. 430-54.
- Kahneman, D. and A. Tversky, 1979. "Prospect Theory: An Analysis of Decisions Under Risk," Econometrica, 47(2), pp. 263-91.
- Kahneman, D. and A. Tversky, 1982a. "The psychology of Preferences," Scientific American, January, pp. 160-80.
- Kahneman, D., P. Slovic, and A. Tversky, 1982b. Judgement Under Uncertainty: Heuristics and Biases, Cambridge, Cambridge University Press.
- Kazan, N. and M.A. Wallach, 1964. Risk Taking: A Study in Cognition and Personality, New York, Holt, Rinehart and Winston.
- Kleindorfer, P.R. and A. Kunreuther, 1983. "Misinformation and Equilibrium in Insurance Markets," in J. Finsinger (ed.), Issues In Pricing and Regulation, New York, Lexington.
- Kneese, A.V., 1962. "Water Pollution: Economic Aspects in Research Needs,"

Washington, D.C., Resources For The Future, Inc.

- Kneese, A.V., 1964. The Economics of Regional Water Quality Management Baltimore, Johns Hopkins University Press.
- Knetsch, J.L., and J.A. Sinden, 1984. "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value," forthcoming, Quarterly Journal of Economics.
- Knetsch, J.L. and R.K. Davis, 1966. "Comparisons of Methods for Recreation Evaluation," in Water Research, A.V. Kneese and S.C. Smith (eds.), Baltimore, Johns Hopkins University Press, pp. 125-142.
- Knight, F.H., 1951. The Economic Organization, New York, Augustus M. Kelley, Inc.
- Kogan, N. and M.A. Wallach, 1964. Risk Taking: A Study in Cognition and Personality, New York, Holt, Rinehart and Winston.
- Kreyszig, E., 1979. Advanced Engineering Mathematics, New York, John Wiley and Sons.
- Krumm, R.J., 1980. "Neighborhood Amenities: An Economic Analysis," Journal of Urban Economics, 7, pp. 208-224.
- Krutilla, J.A., 1967. "Conservation Reconsidered," American Economic Review, 57(4), pp. 777-86.
- Krutilla, J.A., and A.C. Fisher, 1975. "The Economics of Natural Environments: Studies in the Valuation of Commodity and Amenity Resources," Baltimore, Johns Hopkins University Press for Resources for the Future.
- Kunreuther, H.L., 1976, "Limited Knowledge and Insurance Protection," Public Policy, 24(2), pp. 227-61.
- Kunreuther, H.L., R. Ginsberg, L. Miller, P. Sagi, P. Slovic, B. Borkan and N. Katz, 1978. Disaster Insurance Protection: Public Policy Lessons, New York, John Wiley and Sons.
- Lakatos, I., (1970). "Falsification and the Methodology of Scientific Research Programmers," in I. Lakatos and A. Musgrave (eds.), Criticism and the Growth of Knowledge, Cambridge, Cambridge University Press.
- Lancaster, K.J., 1966. "A New Approach to Consumer Theory," Journal of Political Economy, pp. 132-57.
- LaPiere, R.T., 1934. "Attitudes vs. Actions," Social Forces, 13.
- Learner, E.E., 1983. "Let's Take the Con Out of Econometrics," American Economic Review, 73, pp. 31-43.

- Lichtenstein, S. and P. Slovic, 1971. "Reversal of Preferences Between Bids and Choices in Gambling Decisions," Journal of Experimental Psychology, 89, pp. 46-55.
- Lichtenstein, S., and B. Fishchoff, 1978. "Judged Frequency of Lethal Events," Journal of Experimental Psychology, 4(6), pp. 551-78.
- Linegard, C.R., 1943. "Acute and Chronic Studies on Sucrose Octa-acetate by the Oral Method," Bulletin of the National Formulary Committee of the American Pharmaceutical Association, 11, pp. 59-63.
- Linneman, P., 1980 "Some Empirical Results on the Nature of the Hedonic Price Function for Urban Housing Market," Journal of Urban Economics, 8, pp. 47-63.
- Linneman, P., 1981. "The Demand for Residence Site Characteristics," Journal of Urban Economics, 9, pp. 129-149.
- Lipsey, R.L. and K. Lancaster, 1956. "The General Theory of Second Best," Review of Economic Studies, 24(1).
- Little, I.M.D., 1952. "Social Choice and Individual Values," Journal of Political Economy, 60(5).
- Loeb, M., 1977. "Alternative Versions of the Demand-Revealing Process," Public Choice, 29, Special Supplement to Spring Issue.
- Loehman, E., S. Berg, A. Arroyo, R. Hedinger, J. Schwartz, M. Shaw, R. Fahien, V. De, R. Fishe, D. Rio, W. Rossley and A. Green, 1979. "Distributional Analysis of Regional Benefits and Cost of Air Quality Control," Journal of Environmental Economics and Management, 6, pp. 222-43.
- Loomis, J.B., 1983. "A Review of the Suitability of the U.S. Fish and Wildlife Service 1980 National Survey of Fishing, Hunting and Wildlife Associated Recreation in National Assessments and Forest Planning," manuscript. U. S. Forest Service, Ft. Collins.
- Loomis, J.B. and R.G. Walsh, 1982. "An Economic Evaluation of the Benefits and Costs of U. S. Forest Service Wilderness Study Areas in Colorado", Sec. II of Colorado's Wilderness Opportunity, American Wilderness Alliance, Denver.
- Luce, R.D., 1956. "Semiorde and a Theory of Utility Discrimination," Econometrica, pp. 24.
- Machlup, F., 1967. "Theories of the Firm: Marginalist, Behavioral, Managerial," American Economic Review, 57(1).
- Maital, S., 1982. Minds, Markets and Money, New York, Basic Books.
- Majid, J., J.A. Sinden and A. Randall, 1983. "Benefit Evaluation of Increments to Existing Systems of Public Facilities," Land Economics

59(4), pp. 377-392.

Maler, K.G., 1974. Environmental Economics, Baltimore, Johns Hopkins University Press.

Maler, K.G., 1977. "A Note on the Use of Property Values in Estimating Marginal Willingness to Pay for Environmental Quality," Journal of Environmental Economics and Management, 4, pp. 355-369.

Marwell, G. and R.E. Ames, 1981. "Economists Free Ride. Does Anyone Else? Experiments on the Provision of Public Goods. IV." Journal of Public Economics, 15 (3) pp. 295-310.

McCloskey, D.N., 1983. "The Rhetoric of Economics," Journal of Economic Literature, 21, pp. 481-517.

McConnell, K.E. and N.E. Sockstuel, 1984. "Aggregation in Recreation Economics: Issues of Estimation and Measurement," American Journal of Agricultural Economics, 66(5).

McConnell, K.E. and N.E. Sockstuel, 1983. "Welfare Measurement in the Household Production Framework," American Economic Review, 73(4).

McMillan, J., 1979. "The Free-Rider Problem: A Survey," The Economic Record, June, pp. 95-707.

McNeil, B.J., 1982. "On the Elicitation of Preferences For Alternative Therapies," New England Journal of Medicine, 306, pp. 1259-62.

McNeil, B.S. J. Panker, H.C. Cox and A. Tversky, 1982. "Patient Preferences for Alternative Therapies," New England Journal of Medicine, 306, pp. 1259-62.

Medoff, J.L. and K.G. Abraham, 1979. "Can Productivity Capital Differentials Really Explain the Earnings Differentials Associated with Demographic Characteristics? The Case of Experience." Harvard University, Discussion Paper No. 705.

Mellow, W. and H. Sider, 1983. "Accuracy of Response in Labor Market Surveys: Evidence and Implications," Journal of Labor Economics, 1 pp. 331-344.

Mendelsohn, R., 1980, "The Demand and Supply for Characteristics of Goods," mimeo, Department of Economics, University of Washington, Seattle.

Mendelsohn, R. and G.M. Brown, Jr., 1983. "Revealed Preference Approach To Valuing Outdoor Recreation," Natural Resources Journal, 23 (3), pp. 607-18.

Milgrom, P.R. and R.J. Weber, 1982. "A Theory of Auctions and Competitive Bidding," Econometrica, 50.

Miller, G.A., 1956. "The Magical Number Seven, Plus or Minus Two; Some

- Limits on our Capacity for Processing Information" Psychology Review, 63(2), pp. 81-97.
- Miller, G. and C. Plott, 1983. "Revenue Generating Properties of Sealed-Bid Auctions," in V.L. Smith (ed.), Research in Experimental Economics, Vol. 3, Connecticut, JAI Press, Inc.
- Miller, J.R. and M.J. Hay, 1984. Estimating Sub-state Values of Fishing and Hunting," Transactions of the 49th North American Wildlife and Natural Resources Conference, Boston. MA (in press).
- Mitchell, R.C. and R.T. Carson, 1981. An Experiment in Determining Willingness To Pay For National Water Quality Improvements, draft report prepared for the U. S. Environmental Protection Agency, Washington, D.C., Resources for the Future, Inc., June.
- Mitchell, R.C. and R.T. Carson, 1984. "Using Surveys to Value the Benefits of Public Goods: The Contingent Valuation Method," Draft book manuscript, Washington, D.C., Resources for the Future.
- Mitchell, R.C. and R.T. Carson, 1984. "Preliminary Comments on Draft Version of Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Method," Washington, D.C., Resources for the Future, unpublished paper, June 18.
- Mitchell, R.C. and R.T. Carson, forthcoming. "Comment on Option Value: Empirical Evidence From a Case Study of Recreation and Water Quality," Quarterly Journal of Economics.
- Moore, W.C., 1982. "Concept Testing," Journal of Business Research, Vol. 10, pp. 27.
- Morgenstern, O., 1973. On The Accuracy of Economics Observations, 2nd ed., Princeton, Princeton University Press.
- Ness, H.O., 1963. Market Potential For Selected Fee Hunting Enterprises in New Mexico, unpublished M.A. Thesis, New Mexico State University, Las Cruces.
- O'Hanlon, P. and J. Sinden, 1978. "Scope for Valuation of Environmental Goods: Comment," Land Economics, 54(3), pp. 381-87.
- Palfrey, T.R., 1980. "Equilibrium Models of Multiple-Object Auctions," Unpublished Ph.D. dissertation, California Institute of Technology.
- Pashigian, H.P., 1984., "The Effect of Environmental Regulation on Optimal Plant Size and Factor Shares," Journal of Law and Economics, 27 pp. 128.
- Plott, C.R., 1979. "The Application of Laboratory Experimental Methods to Public Choice," in C.S. Russell (ed.), Collective Decision Making, Baltimore, Johns Hopkins University Press.

- Plott, C.R., 1982. "Industrial Organization Theory and Experimental Economics," Journal of Economic Literature, 20.
- Pommerehne, W., S. Schneider and P. Zweifel, 1982. "Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination," American Economic Review, 72, pp. 569-74.
- Pratt, J.W., D. Wise and R. Zeckhauser, 1979. "Price Differences In Almost Competitive Markets," Quarterly Journal Of Economics, 93(2), pp. 189-211.
- Rae, D.A., 1981. Benefits of Improving Visibility at Mesa Verde National Park, Boston, Charles River Associates.
- Rae, D.A., 1983. "The Value to Visitors of Improving Visibility at Mesa Verde and Great Smokey National Parks," in R.D. Rowe and L.G. Chestnut (eds.), Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas, Boulder, Westview Press, pp. 217-34.
- Randall, A., O. Grunwald, S. Johnson, R. Ausness and R. Pagoulatos, 1978a. "Reclaiming Coal Surface Mines in Central Appalachia: A Case Study of the Benefits and Costs," Land Economics, 54(4), pp. 472-89.
- Randall, A., O. Grunwald, A. Pagoulatos, R. Ausness, and S. Johnson, 1978b. Estimating Environmental Damages From Surface Mining of Coal in Appalachia: A Case Study report to the USEPA, EPA-600/2-78-003, Contract No. 68-01-3586, Industrial Environmental Research Laboratory, Office of Research and Development, January.
- Randall, A., J.P. Hoehn and D. Brookshire, 1983. "Contingent Valuation Surveys for Evaluating Environmental Assets," Natural Resources Journal, 23(3), pp. 635-48.
- Randall, A., J.P. Hoehn and G.S. Tolley, 1984. "The Structure of Contingent Markets: Some Experimental Results," forthcoming, Journal of Environmental Economics and Management.
- Randall, A., B. Ives and C. Eastman, 1974. "Bidding Games for Valuation of Aesthetic Environmental Improvements," Journal of Environmental Economics and Management, 1, pp. 132-49.
- Randall, A., B. Ives and C. Eastman, 1974. "Benefits of Abating Aesthetic Environmental Damage from the Four Corners Power Plant, Fruitland, New Mexico," New Mexico State University Agricultural Experiment Station, Bulletin 618.
- Randall, A. and J.R. Stoll, 1980. "Consumer's Surplus in Commodity Space," American Economic Review, 70, pp. 449-454.
- Randall, A. and J.R. Stoll, 1983. "Existence Value in a Total Valuation Framework," Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas, R.D. Rowe and L.G. Boulder (eds.), Colorado, Westview Press.

- Reilly, R., 1982. "Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design," American Economic Review, 72, pp. 576-584.
- Ridker, R., 1967. Economic Costs of Air Pollution, New York, Praeger.
- Robbins, H., 1932. An Essay on the Nature and Significance of Economic Science, London, MacMillan.
- Robertson, L., 1974. "Urban Area Safety Belt Use In Automobiles With Starter Interlock Belt Systems: A Preliminary Report," Washington, D.C., Insurance Institute For Highway Safety.
- Robinson, J., 1962. Economic Philosophy, Garden City, Doubleday.
- Ronan, J., 1973. "Effects of Some Probability Displays on Choices," Organizational Behavior, 9(1), pp. 1-15.
- Rosen, S., 1974. "Hedonic Prices and Implicit Market: Product Differentiation in pure Competition," Journal of Political Economy, pp. 34-55.
- Rothenberg, J., 1961. The Measurement of Social Welfare, Englewood Cliffs, Prentice-Hall, Inc.
- Rowe, R., R.C. d'Arge and D.S. Brookshire, 1980a. "Environmental Preferences and Effluent Charges," Chapter 4 of Progress in Resource Management and Environmental Planning, Vol. 2, T. O'Riordan and K. Turner (eds.), New York, John Wiley and Sons.
- Rowe, R.D., R.C. D'Arge and D.S. Brookshire, 1980s. "An Experiment on the Economic Value of Visibility," Journal of Environmental Economics and Management, 7, pp. 1-19.
- Rowe, R.D. and L.G. Chestnut, 1983. "Valuing Environmental Commodities: Revisited," Land Economics, 59(4), pp. 404-10.
- Samuelson, P.A., 1954. "Pure Theory of Public Expenditure," Review of Economics and Statistics, 36, pp. 387-89.
- Samuelson, P.A., 1955. "Diagrammatic Exposition of a Theory of Public Expenditure," Review of Economics and Statistics, 37.
- Samuelson, P.A., 1958. "Aspects of Public Expenditure Theories," Review of Economics and Statistics, 37.
- Scheffe, H., 1970. "Practical Solutions to the Behrens-Fisher Problem," Journal of the American Statistical Association, 65.
- Schelling, T.C., 1978. Micromotives and Macrobehavior, New York, Norton.
- Scherr, B.A. and E.M. Babb, 1975. "Pricing Public Goods: An Experiment with Two Proposed Pricing Systems," Public Choice, 23, pp. 35-48.

- Schoemaker, P.J.H. and H.C. Kunreuther, 1979. "An Experimental Study of Insurance Decisions," Journal of Risk and Insurance, 46(4), pp. 603-18.
- Schoemaker, P.J.H., 1980. Experiments on Decisions Under Risk: The Expected Utility Hypothesis, Boston, Nijoff.
- Schoemaker, P.J.H., 1982. "The Expected Utility Model: Its Variants, Purposes, Evidence, and Limitations," Journal of Economic Literature, 20, pp. 529-63.
- Schulze, W.D., D.S. Brookshire and T. Sandler, 1981a. "The Social Rate of Discount For Nuclear Waste Storage: Economics or Ethics?" Natural Resources Journal, 21(4), pp. 811-832.
- Schulze, W.D., D.S. Brookshire, E.G. Walther, K. Kelley, M.A. Thayer, R.L. Whitworth, S. Ben-David, W. Malm and J. Molenaar, 1981b. "The Benefits of Preserving Visibility in the National Parklands of the Southwest," Vol. VIII, Methods Development for Environmental Control Benefits Assessment, Final report to the USEPA grant R805059010, Office of Exploratory Research, Office of Research and Development.
- Schulze, W.D., D.S. Brookshire, E.G. Walther, K.K. MacFarland, M.A. Thayer, R.L. Whitworth, S. Ben-David, W. Malm and J. Molenaar, 1983b. "The Economic Benefits of Preserving Visibility in the National Parklands of the Southwest," Natural Resources Journal, 23, pp. 149-73.
- Schulze, W.D., R.G. Cummings, D.S. Brookshire, M.H. Thayer, R.L. Whitworth and M. Rahmatian, 1983a. "Experimental Approaches to Valuing Environmental Commodities: Vol. II," Draft final report for Methods Development in Measuring Benefits of Environmental Improvements, USEPA Grant # CR 808-893-01, July.
- Schulze, W.D., R.C. d'Arge and D.S. Brookshire, 1981c. "Valuing Environmental Commodities: Some Recent Experiments," Land Economics 57, pp. 151-172.
- Schuman, H. and M.P. Johnson, 1976. "Attitudes and Behavior," Annual Review of Sociology, pp. 161-207.
- Scott, A., 1965. "The Valuation of Game Resources: Some Theoretical Aspects," Canadian Fisheries Reports, No. 4.
- Shubik, M., 1975. "Oligopoly Theory, Communication and Information," American Economic Review, 65.
- Simon, H.A., 1955. "A Behavioral Model of Rational Choice," Quarterly Journal of Economics, 69, pp. 174-83.
- Simon, H.A. and A. Newell, 1971. "Human Problem Solving: The State of The Theory in 1970," American Psychology, 26(2), pp. 145-59.

- Simon, H.A., 1979. "Rational Decision Making in Business Organizations," American Economic Review, 69(4), pp. 493-513.
- Sinclair, W.S., 1976. "The Economic and Social Impact of the Kemano II Hydroelectric Project on British Columbia's Fisheries Resources," Vancouver, Fisheries and Marine Service, Department of the Environment.
- Sinden, J. and J. Wyckoff, 1976. "Indifference Mapping: An Empirical Methodology for Economic Evaluation of the Environment," Regional Science and Urban Economics, 6, pp. 81-103.
- Slovic, P. and S.C. Lichtenstein, 1968. "Relative Importance of Probabilities and Payoffs in Risk Taking," Journal of Experimental Psychology, 78(3, Part 2).
- Slovic, P., 1969. "Differential Effects of Real Versus Hypothetical Payoffs on Choices Among Gambles," Journal of Experimental Psychology, 80(3), pp. 434-37.
- Slovic, P., H. Kunreuther and G.F. White, 1974. "Decision Processes, Rationality and Adjustment to Natural Hazards," Chapter 24 in G.F. White (ed.), Natural Hazards, Oxford University Press.
- Slovic, P., A. Tversky, 1974. "Who Accepts Savage's Axioms?" Behavioral Science, 19(6), pp. 368-73.
- Slovic, P., B. Fishhoff and S. Lichtenstein, 1980. "Facts and Fears: Understanding Perceived Risk," pp. 181-216, in R.C. Schwing, and W.A. Albers, Jr. (eds), Societal Risk Assessment: How Safe is Safe Enough? New York, Plenum Press.
- Smith, V.K., 1983. "Option Value: A Conceptual Overview," Southern Economic Journal, 49, pp. 654-668.
- Smith, V.L., 1967. "Experimental Studies of Discrimination vs. Competition in Sealed-Bid Auction Markets," Journal of Business, 40.
- Smith, V.L., 1976. "Experimental Economics: Induced Value Theory," American Economic Review Proceedings, 66.
- Smith, V.L., 1977. "The Principle of Unanimity and Voluntary Consent in Social Choice," Journal of Political Economy, 85, pp. 1125-1139.
- Smith, V.L., 1979a. "Incentive Compatible Experimental Processes for the Provision of Public Goods," in V.L. Smith (ed.), Research in Experimental Economics, Vol. 1, Connecticut, JAI Press, Inc.
- Smith, V.L., 1979b. "An Experimental Comparison Of Three Public Good Decision Mechanisms," Scandinavian Journal of Economics, 81.

- Smith, V.L., 1980. "Experiments with a Decentralized Mechanism for Public Good Decisions," American Economic Review, 70, pp. 584-99.
- Smith, V.L., 1982. "Microeconomic Systems as an Experimental Science," American Economic Review, 72(5), pp. 923-55.
- Sorg, C. and D.S. Brookshire, 1984. "Valuing Increments and Decrements of Wildlife Resources -- Further Evidence," report to U.S. Forest Service Rocky Mountain Forest and House Experiment Station, Fort Collins.
- Starr, C., R. Rudman and C. Whipple, 1976. "Philosophical Basis for Risk Analysis," Annual Review of Energy, 1, pp. 629-662.
- Stigler, G.J., 1950. "The Development of Utility Theory: II," Journal of Political Economy 63(5), PP. 373-396.
- Stigler, G.J. and G.S. Becker, 1977. "De Gustibus Non Est Disputandum," American Economic Review, 67(2), pp. 76-90.
- Strauss, R.P. and G.D. Hughes, 1976. "A New Approach to the Demand for Public Goods," Journal of Public Economics, 6, pp. 191-204.
- Sutherland, R.J. and R.G. Walsh, 1982. "Effect of Distance on the Preservation Value of Water Quality," draft, June.
- Takayama, A., 1982. "On Consumer's Surplus" Economic Letters, Vol. 10, pp. 35-42.
- Thaler, R., 1980. "Toward A Positive Theory of Consumer Choice," Journal of Economic Behavior and Organization, 1, pp. 39-60.
- Thaler, R. and S. Rosen, 1975. "The Value of Saving a Life: Evidence from the Labor Market", pp. 265-98 in N. Terleckys (ed.), Household Production and Consumption, New York, Columbia University Press.
- Thayer, M.A., 1981. "Contingent Valuation Techniques For Assessing Environmental Impacts: Further Evidence," Journal of Environmental Economics and Management, 8, pp. 27-44.
- Thompson, M.E. and K.J. Roberts, forthcoming. "An Empirical Application of the Contingent Valuation Technique to Value Marine Resources," Transactions of the American Fisheries Society.
- Tideman, T. and G. Tullock, 1976. "A New and Superior Process for Making Social Choices," Journal of Political Economy, 84.
- Tolley, G.S., A. Randall, G. Blomquist, R. Fabian, G. Fishelson, A. Frankel, J.P. Hoehn, R. Krumm and E. Mensah, 1984. Establishing and Valuing the Effects of Improved Visibility in the Eastern United States, final report, #80776-01-0, USEPA, Washington, D.C.
- Tversky, A. and D. Kahneman, 1974. "Judgement Under Uncertainty: Heuristics

- and Biases," Science, 185(2), pp. 1124-31.
- Tversky, A. and D. Kahneman, 1981. "The Framing of Decisions and the Rationality of Choice," Science, p. 453.
- Ullman, E. and D. Volk, 1961. "An Operational Model For Predicting Reservoir Attendance and Benefits," Proceedings of the Michigan Academy of Sciences, Department of Economics, University of Michigan, Ann Arbor.
- Van Nostrand's Scientific Encyclopedia (5th ed.) New York, Van Nostrand Reinhold Co.
- Vickery, W., 1961. "Counterspeculation, Auctions and Competitive Sealed Tenders," Journal of Finance, 16, pp. 8-37.
- Vickery, W., 1961. "Auctions, Markets, and Optimal Allocation", in Y. Amihud (ed.), Bidding and Auctioning for Procurement and Allocation New York, New York University Press.
- Walbert, M.S., 1984. "Valuing Policies Which Reduce Environmental Risk: An Assessment of the Contingent Valuation Method," unpublished Ph.D. dissertation, University of New Mexico, Albuquerque.
- Walsh, R.G., 1980. An Economic Evaluation of the General Management Plan for Yosemite National Park, Technical Report No. 19, Water Resources Research Institute, Colorado State University, Fort Collins, March.
- Walsh, R.G., R. Aukerman and R. Milton, 1980. Measuring Benefits and the Economic Value of Water in Recreation on High Country Reservoirs, Completion Report No. 102, Colorado Water Resources Research Institute, Fort Collins, September
- Walsh, R.G., R.K. Ericson and D.J. Arosteguy, 1980. An Application of a Model for Estimating the Recreation Value of Instream Flow, Completion Report No. 101, Colorado Water Resources Research Insititute, Colorado State University, Fort Collins, December.
- Walsh, R.G., R.K. Ericson,, J.R. McKean and R.A. Young, 1978. Recreation Benefits of Water Quality: Rocky Mountain National Park, South Platte River Basin., Colorado, Environmental Resources Center, Technical Report No. 12, Colorado State University, Fort Collins, May.
- Walsh, R.G., D.A. Greenley, R.A. Young, J.R. McKean and A.A. Prato, 1978. Option Values, Preservation Values and Recreational Benefits of Improved Water Quality: A Case Study of the South Platte River Basin, Colorado, Socioeconomic Environmental Studies Series, EPA-600/5-78-001 U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, January.
- Walsh, R.G., G. Keleta and J.P. Olienyk, 1981. Value of Trees to Residential Property Owners with Mountain Pine Beetle and Spruce

Budworm Damage in the Colorado Front Range, Report by the Department of Economics to the Forest Service, USDA, Washington, D.C., October.

Walsh, R.G., N.P. Miller, and L.O. Gillaim, 1982. "Congestion and Willingness to Pay for Expansion of Skiing Capacity," Land Economics, 59 (2), pp. 195-210.

Walsh, R.G. and J.P. Olieniyk, 1981. Recreation Demand Effects of Mountain Pine Beetle Damage to the Quality of Forest Recreation Resources in the Colorado Front Range, Report by the Department of Economics to the Forest Service, USDA, Washington, D.C., October.

Ward, E., 1954. "The Theory of Decision Making," Psychological Bulletin, 51(4), pp. 380-417.

Weinstein, M.C. and R.J. Quinn, 1983. "Psychological Considerations In Valuing Health Risk Reductions," Natural Resources Journal, 23(3), pp. 659-73.

Weisbrod, B.A., 1964. "Collective-Consumption Services of Individual-Consumption Goods," Quarterly Journal of Econmics, 78, pp. 471-77.

Wicksell, K., 1896. "A New Principle of Just Taxation," translated by J.M. Buchanan, in R.A. Musgrave and A.T. Peacock (eds.) Classics in the Theory of Public Finance, New York, St. Martin's Press (1967).

Wilde, L., 1980. "On the Use of Laboratory Experiments in Economics," in J. Pitt (ed.), The Philosophy of Economics, Dordrecht, Reidel.

Willig, R.D., 1976. "Consumer's Surplus Without Apology," American Economic Review, 66(5), pp. 89-97.